SEEDS

Requirements / LOS & Cost Model

Working Paper

New Year's Draft

January 16, 2002

G. Hunolt, SGT, Inc.

NOTE for Fifth (January 16) version:

The fourth version of the working paper provided to ESDIS SOO was the December 21 version.

The Roadmap & Notes on this Version (Sect 1.2) will highlight changes since the December 21 version.

The next version provided to ESDIS SOO will highlight changes made since this version.

This is one of them there pages left mostly blank!

Outline

1 Introduction

This working paper describes the SEEDS (Strategic Evolution of Earth Science Data Systems) cost estimation model and coupled requirements sets developed to support the SEEDS Formulation team in estimating the life cycle costs of future ESE data service providers and supporting systems, where 'data service provider' is used as a generic term for any data/information related activity. (As of late November, 2001, SEEDS replaces the term 'NewDISS' under which the Formulation Team had begun work. The new term is intended to emphasize the Earth Science Enterprise's (ESE's) evolutionary approach. The term 'NewDISS' will be retained when it refers to NewDISS documents that predate the change in terminology.) This working paper is intended to serve as a vehicle for coordinating work on the project, obtaining feedback and guidance from ESDIS SOO, and as the embryo of one or more reports that will be produced as the task proceeds.

As a working paper, each version that appears represents a snapshot in time, with the work in various stages of completion; readers should expect loose ends and inconsistencies especially in the early stages of the project. As work progresses the content (and sometimes the organization) of the working paper will change reflecting progress made, responses to feedback and guidance received, etc.

The introduction covers task objectives in Section 1.1, introduces the following sections of the paper in Section 1.2, reporting on the current status of each section of the paper (i.e. the work covered by that section) and concludes with a list of work to be done in Section 1.3.

1.1 Objectives

A key facet of the SEEDS Formulation study will be to establish the minimum levels of service that ESE data service providers will be required to provide for the user community and to provide the capability to estimate costs for ESE data service providers to provide that level of service. The ultimate objective of this study is to provide the SEEDS Formulation Team with a capability to estimate the cost for various system architectures and mission profiles. Successful development of a life cycle cost estimation capability will be dependent on an accurate assessment of the levels of services needed from ESE data service providers.

"Levels of service" will be associated with certain functional requirements, describing different degrees of performance with which the requirement would be met. For example, a functional requirement might be: "The data service provider shall distribute data and products to users on media". Accompanying this requirement might be descriptions of quantitatively distinct levels of service, such as "delivery on media shall be provided within one working day of receipt of a data request", "delivery on media shall be provided within two calendar weeks of receipt of a data request", and "delivery on media shall be provided within one calendar month of receipt of a data request". Which level of service would be most appropriate ('recommended') or acceptable ('minimum') for a particular ESE data service provider would depend on its particular mission and the needs of its users.

The first objective of this study is to assist the Formulation Team in establishing the minimum (and recommended) levels of service (LOS) for ESE data service providers. These LOS will be refined in a bottoms-up manner through community workshops of potential providers and users of ESE data service providers.

The second objective of this task is develop a suite of costs estimation tools that will enable the FT to estimate the cost impact for various architecture trades, provide NASA Headquarters with estimates of the costs for implementing varying ESE mission profiles and implementation options, and packaging the cost estimation tool kit for use by Earth Science Enterprise scientists responding to new mission opportunities in order for them to estimate the costs for developing and operating the science data ground system for their proposed mission.

The purpose of working on these two objectives together is to ensure that the cost estimation process is tied to a reasonable requirements / levels of service set.

1.2 Roadmap to the Paper and Notes / What's New for Version 5, January 16, 2001

Note: What is new in the January 16, 2002 version (Version 5) is indicated in italics below.

A global change in terminology is the adoption of the term 'data service provider' as the generic name for an ESE activity that provides any form of data and/or information management and user services, replacing the term 'data center', which will be used only in the more conventional sense as a type of data service provider.

Section 2 discusses data service provider cost estimation, including the cost estimation tool and the output to be provided by it. An in-depth understanding of what is required by the SEEDS Formulation Team from the cost estimation tool will guide its development, so this section will dissect the cost estimation objective given by the Formulation Team and lay out what the study will provide to meet that objective. The section discusses what a cost estimate for implementing and operating an ESE data service provider would include, and the use of cost estimates in considering ESE architecture trades. The section includes an initial draft template of an output cost estimation table for review, comment, and revision. It also outlines an initial operations concept, i.e. an initial scenario illustrating how the cost estimation tool would be used and clarifies how the cost estimation tool would support ESE architecture trades. *No change in the January 16 Version.*

Section 3 presents the technical approach, including some related notes and assumptions, and the approach to be taken to the requirements analysis and cost model development, consistent with the task plan submitted in draft to ESDIS SOO on October 19. A summary section, 3.1.5, includes a figure (presented at the NewDISS Retreat, November 2000) illustrating the relationship between the reference model, requirements / levels of service, ESE data service provider types, and cost estimation. *For the January 16 Version, some redundant material was trimmed out and some points were clarified. The figure is deleted from this version - for some reason it becomes hopelessly garbled when the document is sent anywhere, a mystery yet to be resolved.*

Section 4 discusses the general data service provider reference model, the functional areas it includes, and the parameters associated with each functional area, and indicates their association with the requirements / levels of service in Section 5. Section 4.2 explains the relationship between the reference model parameters and the requirements. *For the January 16 Version, Section 4 has been completely reorganized and redone. A more straightforward approach to identifying parameters has been adopted, and a number of clarifications to parameter definitions, additions/deletions of parameters were made. The mapping of the requirements in Section 5 and parameters in Section 4 has been begun but remains incomplete.*

Section 5 discusses the requirements analysis, and presents the general ESE data service provider requirements and levels of service, and their mapping to the reference model. The requirements and associated levels of service reflect the view of the ESE data service provider types presented in Section 6 and the distinctions between the data service provider types and the services they would provide. The USGCRP report, CES report, and Level 2 EOSDIS Version 0 requirements / levels of service (see references) were reviewed for guidance on requirements / levels of service. *For the January 16 version, some annotations about computational approaches have been removed to keep the focus of the section on requirements/LOS.*

Section 6 discusses the ESE data service provider types, based on material taken from the October, 2000, NewDISS draft concept paper. The descriptions of the data service provider types include a reference model functional area by functional area description of each one. The intent is to not only describe the types but highlight differences between them. Section 6 includes tables showing the mapping of the general data service provider requirements / levels of service to the seven ESE data service provider types. *No change for the January 16 Version.*

Section 7 discusses the "ESE EOSDIS and other Data and Information-Related Assets". The information is organized by the current classifications of these activities (e.g. DAACs, ESIPs, RESACs, Pathfinders, NASA Space Science, Non-NASA, etc.). A brief description of each activity is included, along with an assessment of

where that activity seems to map in terms of the ESE data service provider types. The compilation of this information is in progress, so the section is incomplete. A concluding subsection, yet to be added, will present a summary table showing the mapping of the existing activities to the ESE data service provider types, will indicate those that seem the best prospects to either be added to the cost model's 'comparables' database or held aside to serve as independent test cases. *No changes in the January 16 Version.*

Section 8 will document the cost estimation relationships that are incorporated into the model. The barely embryonic section includes only an explanation of how the section is expected to develop, along with notes on an approach to a simple 'bare bones' proto-prototype starter cost model that might be a good idea to do to shake down some ideas. *No change in the January 16 Version.*

1.3 Work To be Done, With Status as of January 16

This section lists items to be done. As they are 'completed', or become 'in progress', this will be indicated with a date. For this study, 'completed' really means 'completed for now' - changes as the study progresses or in response to feedback will occur through the life of the study. When an item is tied to a due date, that will be indicated.

Items, and their status as of January 16:

1. Requirements:

    1) Ensure requirements / levels of service are coupled with the model development. This is in progress as of January 16.

    2) Refine requirements / levels of service set based on the community workshop that will include review of requirements / levels of service scheduled for February 5-7, 2002.

    3) Produce final requirements / levels of service set report. Due March 31, 2002.

2. Extend the general reference model to encompass the full range of ESE data service provider functions (items will be added as needed and will be deleted from the list after they've been completed a while):

    1) Add the time dimension to what is now a snapshot in time model. Future as of January 16.

    2) Account for network connections between ESE data service providers. Only placeholders as of January 16.

3. Map the information collected on selected data service providers during the Best Practices / Benchmark study to the extended reference model to begin building the model's 'comparables' database;

    1) Collect needed information from existing data service providers. Collection is future as of January 16.

4. Once there is a good enough database, develop cost estimating relationships! Future as of January 16.

5. See what we might be able to do with COTS cost estimation models (in progress as of January 16):

    1) Decide on how we can use, if we can use, COTS cost estimation tools.

6. Examine Bruce Barkstrom's user model, see if we can make use of it. In progress as of January 16.

7. Begin testing a prototype executable cost estimation tool. Due to start testing April 1, 2002.

8. Release first version of the cost estimation tool for evaluation. Due June 1, 2002.

9. Obtain feedback on prototype from community workshop. Workshop scheduled for June, 2002.

10. Provide Final Report on Initial Phase, due June 30, 2002.

2 Data Service Provider Cost Estimation

Section 2 discusses the cost estimation output to be provided by the data service provider cost by analogy (i.e. by comparison to comparable data service providers) model discussed in the following sections of the paper. An understanding of what is required from the model will guide its development.

As noted above, the cost estimation tool is needed to enable the SEEDS Formulation Team to estimate the cost impact for various architecture trades, and to provide NASA Headquarters with estimates of the costs for implementing varying ESE mission profiles and implementation options. The Formulation Team also requires that the tool be packaged so that it can be provided to ESE scientists for their use in estimating the costs for developing and operating the science data ground system for their proposed mission.

The remainder of this section will examine the particulars of this objective and what the cost estimation tool must be able to do to meet it.

2.1 Cost Estimation Tool

In this discussion and throughout the working paper, the term "cost estimation tool" will be used to mean the cost estimation model packaged in a useable form, i.e. provided in a package that can be started up, can receive a set of inputs, run, and produce a set of outputs. The early goal of the study is to provide the tool in as readily useable form as possible, for example as an Excel spreadsheet workbook that could loaded and used on any PC or Macintosh platform equipped with Excel.

In parallel with the effort described below, the study will be examining COTS cost estimating tools (e.g. parametric cost models) to see if one or more of these might be better for certain aspects of costing than the cost model to be developed during the study, or valuable for use in producing alternative cost estimates for some or all aspects of costing.

At least some COTS tools can be integrated with other software such as Excel, and so it may be possible to deliver a cost estimation tool with an integrated COTS component. In any case, the most practical approach will be taken to facilitating the use of any selected COTS tool in conjunction with the model developed by the study.

2.2 ESE Data Service Provider Cost Estimates

The cost estimation tool must provide estimates of implementation and operating costs of individual ESE data service providers (otherwise referred to as ESE components) over a specified lifetime. There will be a number of different types of ESE data service providers (see Section 6 below), each having a particular set of functions to perform. For the purpose of cost estimation, it is assumed that each ESE data service provider must meet a set of requirements appropriate to its particular role within the ESE program. These would be a generic set for its ESE data service provider type rendered specific by the particular requirements of its specific mission, including the specific levels of service it must provide for certain functional requirements.

The implementation period costs of each ESE data service provider must include hardware purchase, custom software development and COTS purchase, integration and test costs, and facility preparation costs. The operating period costs of each ESE data service provider must include hardware maintenance, continuing COTS support, sustaining engineering, operations, recurring facility costs, supplies such as storage and distribution media, and must allow for the possibility of 'technology refresh'. Implementation and operations period staff costs must allow for reasonable management staffing, and labor rates must allow for overhead and inflation adjustments.

Tables 1, 2, and 3 below are together an initial example of what the cost estimation output might look like. The categories would be defined in detail below. Note that the actual number of years for which costs would be estimated (shown as seven in the example) would be selectable as appropriate for actual cases.

Table 1 - Initial Draft Sample of Cost Estimation Output - Implementation Period Costs

ESE Data Service Provider: [Name], Data Service Provider Type: [Type Identifier]								
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Total
Estimated Implementation Costs								
Management Staff, FTE								
Management Staff Cost								
Development Staff, FTE								
Development Staff Cost								
Hardware Purchase								
COTS Software Purchase / License								
Facility Preparation								
Total Implementation FTE								
Total Implementation Cost								

The cost estimate example shown above and below contains some FTE lines that would be generated by the model in the process of producing the cost estimate. Other such lines could be added, such as SLOC to be developed and maintained.

There are a variety of "workload" parameters that could be presented in conjunction with the cost estimate. These could include those that characterize the mission of the ESE data service provider, which would have been provided as input to the cost estimation, such as flight mission to be supported, input data streams, output product streams, etc., appropriate to the type and particular mission of the data service provider.

Table 2 - Initial Draft Sample of Cost Estimation Output - Operations Costs

ESE Data Service Provider: [Name], Data Service Provider Type: [Type Identifier]								
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Total
Estimated Operations Costs								
Management Staff FTE								
Management Staff Cost								
Technical Coordination Staff FTE								
Technical Coordination Staff Cost								
Sustaining Engineering FTE								
Sustaining Engineering Cost								
Engineering Support FTE								
Engineering Support Cost								
Operations Staff FTE								
Operations Staff Cost								
Development FTE								
Development Staff Cost								
Recurring Network / Comm Cost								
Recurring COTS S/W Cost								
Hardware Purchase Cost								
Hardware Maintenance Cost								
Supplies Cost								
Recurring Facility Cost								
Total Operations FTE								
Total Operations Cost								

Table 3 - Initial Draft Sample of Cost Estimation Output - Total FTE and Costs

ESE Data Service Provider: [Name], Data Service Provider Type: [Type Identifier]								
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Total
Estimated Total FTE								
Estimated Total Cost								

All of the parameters shown in the tables above are defined in Section 4, as well as the required input parameters and all of the internal parameters used by the cost estimation model.

2.3 Cost Estimation Operations Concept

This section outlines briefly how the cost estimation tool would be used.

Assume a user wishes to obtain a cost estimate for a new ESE data service provider for which a specific mission has been defined. The following steps would be taken:

a. The user determine the new data service provider's ESE data service provider type;

b. The user select the requirements template for that type;

c. The user creates specific requirements for the new ESE data service provider by replacing the placeholders in the appropriate requirements template with actual quantitative data derived from the data service provider's mission responsibilities.

d. The user provides externality assumptions such as an inflation rate;

e. The model executes to produce the life cycle cost estimate for the ESE data service provider.

The cost estimate will be a life cycle cost including year by year development and sustaining engineering costs and operations staffing and costs projected over a number of years.

(A future appendix could include an end-to-end walk through of an example of the process described above.)

2.4 Architecture Trades

An objective of the cost estimation tool is to support the SEEDS Formulation Team in considering architecture trades. This section addresses what is needed for this purpose.

An overall ESE data system architecture is a collection of ESE data service providers and their interconnections. An architecture would be developed by analyzing the complete set of ESE program requirements for missions to be supported, science and applications efforts to be supported, data holdings to be maintained, the needs of various user communities within and without the ESE program, etc., and determining a set of ESE data service providers (and their required interconnections) able to meet the various ESE mission requirements. There will be multiple possible architectures, i.e. multiple possible configurations of ESE data service providers, that will meet a given set of ESE mission requirements, and hence the need for the ability to analyze trades between them to select the architecture to implement.

ESE mission requirements and hence the needed set of ESE data service providers will vary over time, given factors such as the launch dates for flight missions, the phasing of efforts in the science program, possibly the rotation of data into non-NASA long term archives, etc. As a result there will not only be multiple possible architectures at any given time, but also multiple possible paths for the evolution of the overall ESE architecture as time goes on. That is to say there will be multiple, time varying configurations of ESE data service providers that comprise possible ESE architectures. These may represent different approaches to consolidation of new mission requirements at ESE data service providers, different assignments of level of service requirements, different consolidations of new mission requirements with ongoing (i.e. EOSDIS) mission requirements, etc., and combinations of the above.

For each possible architecture, the cost estimation tool would generate cost estimates for the individual components and allow a roll-up of these to accumulate to an overall estimated cost for the candidate architecture. This would then be one factor taken into account in consideration of the trades between possible architectures (along with complexity, robustness, etc., etc.).

2.5 Portable Version of the Cost Estimation Tool

In addition to their own use of the cost estimation tool, the SEEDS Formulation Team intends to provide the tool to ESE scientists (and possibly applications groups) who are planning the development and operation of a science ground data system for their proposed mission. As noted in Section 2.1, the early goal is to deliver the tool in a form able to be run on any PC or Macintosh platform, e.g. as an Excel workbook application. The precise form of the cost estimation tool is TBD as of now, but might include an integrated COTS tool, or the package might include both an Excel based tool and a separate COTS tool or tools. In any case, the package will be documented as a whole, with a users' guide covering the entire package, supplemented by COTS

documentation as needed. The follow-on effort to the initial phase of this study (i.e. after June, 2002) will include user support for ESE scientists (or applications groups) who make use of the cost estimation tool.

3 Technical Approach

This section outlines the technical approach taken to meeting the objectives stated above. The two objectives are inseparably coupled; costs must be driven by requirements, and so the cost estimation tools must be based on a model that maps directly to the requirements set. For this reason the effort consists of two parallel and intertwined tracks that will merge in the final product. The first track is a requirements analysis, and the second track is development of the cost model. The work on the two tracks is closely coupled, as the requirements must map to the same framework as the costs, and the concept of a data service provider reference model (developed initially for the Best Practices / Benchmark Study) will be used as described below to provide the common framework.

The cost estimation model will be based on a 'comparables' or 'cost by analogy' method; it will estimate costs using cost estimating relationships derived from a number of existing data service providers that are functionally comparable to the different types of ESE data service providers.

3.1 Notes and Assumptions on the Technical Approach

This section contains notes that are background for the discussion of the technical approach to the requirements analysis and model development that follow below in Sections 3.2 and 3.3.

3.1.1 Requirements and Cost Modeling

The requirements developed by this study are intended to support the cost model, and not to serve as the complete definition of the requirements side of a contract between the SEEDS program office and ESE data service providers, or as a basis for procurements. For those purposes this requirements set will be incomplete, perhaps in some cases too detailed, perhaps in other cases not detailed enough. These requirements will be 'end-to-end' in that they will encompass all significant elements of cost, and will be directly and explicitly traceable to cost.

3.1.2 Data Service Providers, the Reference Model, and Requirements

The term 'data service provider' is used herein as a broad, generic term for a site that performs all or a subset of the functions defined in the general data service provider reference model. Many well known actual data centers such as the Distributed Active Archive Centers or the NOAA data centers will perform a subset of the general list of functions, while some sites described as 'data service providers' for this study, e.g. MODAPS (as a sample of a science team processing facility that does not perform archive or general user distribution), are different in function from many well known data centers but fit within the framework of the data service provider reference model.

The general data service provider reference model will have subsets corresponding to the tentatively defined ESE data service provider types. (This approach has the advantage of allowing the future definition of additional data service provider types, or variations of the types defined herein, i.e. other possible subsets, within the same general framework.) In the same manner, a set of requirements / levels of service will be defined corresponding to the general data service provider, and this general set which will have subsets corresponding to each of the defined data service provider types. The data service provider types will provide a basis for 'binning' like data service providers in the cost model data base, allowing the cost estimate for a future data service provider to be based on similar existing cases.

As noted, a set of requirements will be developed for each defined data service provider type. Such a requirements set for a data service provider type will be a template containing placeholders for quantitative parameters that will be defined for a specific instance of a data service provider of that type. For example, a requirement in the template might be that "the data service provider shall provide an archive capacity of

[number TB]". A data service provider of a type that would include providing an archive would have that item in its template. If the mission of the data service provider required that it archive certain data streams and generated products that would accumulate to a total volume of 100 TB, then that value would be inserted into the template, with the result being a specific requirement for that data service provider (i.e., "the data service provider shall provide an archive capacity of 100 TB") that could then be used in the process of generating a cost estimate for the data service provider.

3.1.3 COTS Cost Estimation Tools

In parallel with the above, the possible use of COTS cost estimation tools will be explored, for example for software development, to check the results of the comparables model described above, or perhaps to replace it for aspects of the cost modeling where a COTS tool proves to be superior in tests against the independent cases. This requires an examination and evaluation of the available COTS tools, selecting the most promising for test, and exercising them. While it is likely that there's no existing end-to-end data service provider model, COCOMO-2 may be a good model for software development, etc.

3.1.4 User Model

The life cycle cost model will need to project user demand for a data service provider's services over a period of time. In addition to data service provider history information, the effort will include an examination of existing user models including Bruce Barkstrom's.

3.1.5 Summary

The figure below summarizes the discussion of the relationship between the general data service provider reference model, requirements / levels of service templates, data service provider types as subsets of the general model, the 'comparables' data base of information about existing data service providers and activities mapped to the reference model, and the cost model with its cost estimating relationships derived from the 'comparables' data base. The diagram shows user input to the cost model, providing the specific information about a particular ESE data service provider, and the output being cost projections for that data service provider.

FIGURE TO BE PROVIDED.

It can be seen from the figure that the data service provider reference model, essentially an array of functions vs parameters (metrics) that apply to the functions, is the core concept. The requirements and levels of service are mapped to the functions as templates with placeholders for quantitative values; the placeholders are replaced with actual values when the template is applied to a particular data service provider case. The 'comparables' database contains, for each of a number of real data service providers or activities, sets of real values for the same parameters for those functions performed by each data service provider.

3.2 Technical Approach - Requirements Analysis

The requirements analysis would proceed as follows:

a. Review existing NewDISS and ESE program documents, and incorporate the draft "NewDISS Level 0 Requirements, September 2001", as a high level programmatic framework. Review the EOSDIS Level 2 Requirements for Version 0 as a reference to a previously defined set of requirements and levels of service that could be a source for the current effort. Review USGCRP and CES reports (see references) for additional input on requirements / levels of service.

b. Develop of an initial set of requirements templates with levels of service as applicable, based on an initial version of the data service provider reference model and data service provider subsets, consistent with program documents, for review, revision as needed, and approval as a starting point by the SEEDS Formulation team.

c. Use a community workshop to get user feedback on, and input into, the requirements and levels of service definitions. Produce updates to the requirements templates set, for review, revision as needed, and approval by the SEEDS Formulation team.

d. Produce a final requirements templates set for review, revision as needed, and approval by the SEEDS Formulation team, and produce a final report on the requirements analysis.

At each step, as changes to the requirements sets are approved by the Formulation team, ensure that the requirements changes are reflected back into the data service provider reference model.

3.3 Technical Approach - Cost Estimation Model Development

The general approach is to draw on the data service provider reference model concept developed for the Best Practices / Benchmark study and to develop a cost estimation model that estimates the cost of an ESE data service provider based on the actual costs of comparable data service providers, a "cost estimation by analogy" methodology. Information about other ESE or outside data service providers will be collected to provide the best possible basis for comparison.

In parallel, the use of COTS cost estimation tools will be explored, for example for software development, to check the results of the comparables model described above, or perhaps to replace it for aspects of the cost modeling where a COTS tool proves to be superior in tests against the independent cases. This requires an examination and evaluation of the available COTS tools, selecting the most promising for test, and exercising them.

The life cycle cost model will project user demand for a data service provider's services over a period of time. In addition to data service provider history information, the effort will include an examination of existing user models including Bruce Barkstrom's.

The model development will proceed with the following steps:

a. Define, and refine based on feedback from ESDIS and the Formulation Team, the content of a data service provider cost estimate; i.e. what elements of cost at what level of detail with what supporting information are required as the output product from the cost estimation tool. The further development of the cost model would be guided by the results that the model must produce, allowing for the fact that this will change as the effort proceeds.

b. Survey available COTS cost estimation tools, evaluate and test the tools that seem most likely to be useful for this study, and produce a report summarizing the results of the survey and recommending tool(s) to be used further in the study.

c. Obtain and examine the Bruce Barkstrom user model and any other user model that might be useful for this study.

d. Extend the existing Best Practices / Benchmark study reference model to encompass the full range of data service provider functions, refine the original list of model parameters, add implementation, parameters necessary for estimation of cost, etc.

e. Derive subsets of the general reference model to correspond to the data service provider types. These subsets will include the functional areas and metrics appropriate to each data service provider type.

f. Map the information collected on selected data service providers during the Best Practices / Benchmark study to the extended reference model to begin to build the model's 'comparables' information set;

g. Identify additional data service providers to be added to the model's information set. Draw from DAACs not included in the Best Practices / Benchmark study, SIPS in addition to MODAPS and TSDIS (e.g. SeaWiFS, AMSR-E), selected ESIPS such as GHRC. Consult with ESDIS to arrive at a list of candidates. Reserve some data service provider cases for use as independent test cases for the cost model.

h. Collect the additional or update information and add to the model's information set (i.e., as was done for the Best Practices / Benchmark study, map data service provider information to the reference model's common set of metrics).

Note that steps b and c can run in parallel with a, d, etc. Also, steps g and h can run in parallel with d, e, and f provided that information collected early on can be supplemented as completion of steps d, e, and f identify gaps in the initial collection.

i. Use the mapped data service provider information to construct relationships (for each data service provider type, within each functional area) relating actual data service provider staffing and costs and known development effort and workload performed, etc. These relationships are currently TBD but could include linear regression equations and the like. Probable errors of estimate will also be derived for each relationship.

j. Test the model by inputting information for the independent test cases and determining the degree to which model is able to correctly calculate staffing, costs, etc. Test the COTS cost estimation tools, to determine which should be incorporated into the model or used in conjunction with the model to give the best possible overall result. Also consider incorporation of the Bruce Barkstrom or other externally developed user model.

k. Obtain community feedback on the prototype cost estimation model.

l. Release a life-cycle cost model Version 0 that incorporates initial user feedback; continue to obtain and incorporate community feedback by presenting study results and providing prototype models for hands-on peer review.

m. Provide a final report, and provide cost estimates for ESE data service providers as needed.

4 General Data Service Provider Reference Model

This section describes the Data Service Provider Reference Model, a model of a data service provider in terms of a set of functional areas that, taken together, comprise the range of functions that a data center performs and the areas of cost that must be considered by the cost estimation model. Section 4.1 describes the reference model's functional areas, Section 4.2 describes how the reference model parameters are organized, how they are used by the cost estimation model, and their relation to the requirements / levels of service addressed in Section 5. Sections 4.3 contains the parameter descriptions, and Sections 4.4 and 4.5 contain parameter lists (cost model output, user input, and control parameters).

4.1 Reference Model Functional Areas

The first step is the outlining of a simple reference model of a generic ESE data service provider in terms of a set of functional areas that taken together comprise the full range of functions that a data service provider might perform and implementation and operating costs it would incur. The different types of data service providers would function within all or some of these areas. In each functional area, the cost estimation model will consider implementation and operations costs. Some of the areas are not strictly speaking "functional" in nature (such as 'facility / infrastructure') but are needed to ensure that all significant cost areas are included. The list of functional areas can be taken in toto as a summary of the areas of cost included in the model.

The following are working definitions of the functional areas that make up the data service provider reference model:

- **Ingest**—the process of receiving, reading, quality checking, cataloging, of incoming data to the point of insertion into the archive. Ingest can be manual or electronic with manual steps involved in quality checking, etc. Incoming data can be received from external sources or internally generated. Ingest can include format conversion, metadata extraction, or other preparation of incoming data for archive or use within the data service provider.
- **Processing**—the generation and quality checking of new derived data products from data or products that have been ingested, or previously generated, generally on a routine, operational basis. Processing includes process control (production planning, scheduling, monitoring, etc.) as well as product generation per se. The data service provider may receive the software that embodies product generation algorithms from outside developers (e.g. some Terra instrument teams for the DAACs currently) who are responsible for the initial delivery and for delivering updated versions. Where quality, especially science quality, of products remains the responsibility of an outside developer, processing includes quality checking by the science software developer. Support provided by the data service provider for integration and test of this 'science software' is included as an activity under processing. In cases where a data service provider develops algorithm software, that effort (i.e. development, integration, and test) is included under Development.
- **Documentation**---the development (or upgrading of received) data and product documentation to meet adopted standards, including catalog information (metadata), user guides, etc., through consultation with data providers, algorithm developers, flight projects, etc.
- **Archive**—the insertion of data into archive storage, and management, handling and preservation of data, metadata, and documentation within a data service provider's archive. Inserted data can include data ingested from sources external to the site, or data/products generated on-site. Handling and preservation include quality screening of data entering and exiting the archive, quality screening of archive media, backups, and accomplishing migrations from one type of media to another. Insertion into the archive can be electronic or manual (e.g. hanging tapes on a rack or popping them into a robotic silo).
- **Distribution**—providing access to catalog information and a search and order capability to users, receiving user requests for data, fetching the requested data from the archive, performing any subsetting,

reformatting / format conversion, or packaging, and providing the end product to the user by electronic means or on physical media. Catalog search and order can include providing local user interface and capability and/or providing an interface to a broader based, cross-site search and order capability (e.g. DAACs supporting search and order via the EOS Data Gateway).

- **User Support**—user support provided in direct contact with users by user support staff, including responding to queries, taking of orders, etc.
- **Instrument / Mission Operations** - monitoring instrument and spacecraft performance, generating instrument and spacecraft commands, and event scheduling (using NASA or other appropriate operational mission management services).
- **Sustaining Engineering** - Maintenance and enhancement of custom applications software (including any science software embodying processing algorithms developed by the site).
- **Engineering Support** - Some or all of the following as applicable at a particular site: systems engineering, test engineering, configuration management, coordination of hardware maintenance by vendors, COTS procurement, installation of COTS upgrades, system administration, database administration, network/communications engineering. Engineering support is internal, directed toward the internal operation of the data service provider.
- **Technical Coordination** - Includes participation in SEEDS system level processes, including coordination on standards and interfaces, work on common metrics, overall architecture, etc. Technical coordination, which by its nature includes engineering, is directed outward, supporting the data service provider as one element of a system of cooperating centers.
- **Implementation---**Includes development of, and making operational, the data and information system capabilities required by the data service provider to perform its mission, including design and implementation of the data system (hardware and system software) and applications software. Implementation can recur during the operating period as systems are expanded or replaced. In some cases applications software will include product generation software embodying science algorithms. Development can include development of software tools for use by users to unpack, subset, or otherwise manipulate products provided by the data service provider.
- **Management** - Includes management and administration at the data service provider level ("front office") and direct management of functional areas, and internal support: some or all of the following as applicable at a particular site: logistics, supplies, facilities, security management, property inventory and management, facility management.
- **Facility / Infrastructure** - Includes a variety of non-staff cost factors such as supplies, facility lease and utility costs and other similar overhead costs, hardware maintenance, COTS licenses, etc.

4.2 Reference Model Parameters

This section sets the stage for the description and definition of reference model parameters that follows in sections 4.3, 4.4, and 4.5. The reference model parameters are a standard set of parameters that includes some that cover a data service provider as a whole and some that are mapped to the model's functional areas as they apply (i.e., not all parameters are applicable to all functional areas).

The scope of the parameters spans implementation and operations, year by year over the specified lifecycle of the data service provider, and include cost elements as well as workload factors and high level system configuration information.

The implementation and operations parameters will be broken down into outputs to be provided by the model, internal (derived) parameters used by the model, and inputs required by the model.

The cost estimation relationships to be used by the model will be derived from information describing actual date centers comparable to future ESE data service providers. As was done for the Best Practices / Benchmark Study, raw information received from the data service providers will be mapped to the standard reference model parameter set to build the model's database, so that the model's database will contain the same set of

output, input, and derived internal parameters covering implementation and operation as will be used for cost estimation. This is necessary, since the model database will be used to derive the cost estimation relationships that allow estimation of the outputs given the inputs for independent cases (i.e. testing against independent data for an actual data service provider and use of the model to estimate the costs for a putative new ESE data service provider).

Implementation includes capital and staff costs associated with developing, implementing, integrating and testing the data service provider's data and information system, and facility start-up / preparation costs. Implementation is assumed to be spread over a specified number of years. Implementation can overlap the start of operations. Implementation can also recur during the operating period, e.g. allowing for 'technology refresh'.

Operation includes hardware maintenance, sustaining engineering, operations staff, supplies (e.g. storage and archive media), recurring facility costs, etc.

Section 4.3 discusses the reference model parameters. These are grouped by the functional areas described in Section 4.1 above. Within each functional area group, the parameters are sorted into as internal derived parameters used by the model, and input parameters that must be provided by a user of the cost estimation tool, and output parameters, i.e. required outputs from the cost estimation tool.

Section 4.4 contains a list of the cost estimation model output parameters, and Section 4.5 contains a list of the user input parameters required to run the cost estimation model. Both lists are drawn from the parameters defined in Section 4.3.

Section 5 describes the requirements / levels of service template for the reference data service provider described by the general reference data service provider model. The cost estimation input parameters in Section 4 are mapped to requirements in Section 5 in order to ensure that the cost estimate is driven by real requirements. For example, assume that a data service provider will have to ingest a certain volume of data in order to meet its mission responsibilities. The volume of data it must ingest will affect its implementation and operation cost. Therefore in Section 5 there will be a requirement that the data service provider shall ingest a certain volume (or that the data service provider shall ingest a list of data streams whose volume totals to a certain volume), and in Section 4 there will be a corresponding operational workload parameter, the volume of data to be ingested, that is needed for cost estimation. The mapping between the requirements and the reference model parameters will be indicated in Section 4 and Section 5 by appropriate annotation of the parameters and requirements.

4.3 Reference Model Parameter Definitions

This section contains a master list of the data service provider reference model parameters and their definitions. The list is grouped by the functional areas described in Section 4.1, followed by facility / infrastructure parameters and data service provider level parameters (some of which are roll-ups from the preceding functional areas). Within each functional area, the parameters will be sorted between internally computed parameters, parameters provided as user input when executing the cost estimation model, and cost estimation model output parameters.

Information included about each parameter is:

a. Parameter Name;

b. Parameter Definition;

c. Reference to Requirements / Levels of Service (provided in Section 5).

4.3.1 Ingest

These parameters describe or relate to the ingest of data and products into the data service provider from external sources / providers.

4.3.3.1 Internal Computed Parameters

1. **Total Ingest FTE.** The total estimated annual FTE effort for the Ingest functional area, including any effort in addition to actual operational effort.
2. **Ingest Management FTE.** Includes direct management associated with the Ingest functional area. Computed from technical and operations staffing.
3. **Ingest Technical FTE.** Includes ingest technical staff exclusive of direct operations staff.
4. **Ingest Ops FTE**. The estimated annual FTE effort for direct operational activity (e.g. computer operators, ingest technicians).
5. **Ingest Volume/Yr.** The annual volume of data and/or products that are ingested by the site. {Maps to requirement 5.4 a}
6. **Ingest Volume/Yr per FTE.** The annual volume divided by the total staff effort for the Ingest functional area.
7. **Ingest Volume/Yr per Ops FTE.** The annual volume divided by the direct operations staff effort for the Ingest functional area.
8. **Product Types Ingested/Yr.** The annual number of different product types ingested (i.e. data streams ingested) from external sources by the site. {Maps to requirement 5.4 a}
9. **Product Ingest Formats/Yr.** The number of distinct different product or data formats handled by the Ingest functional area.
10. **Products Ingested/Yr.** The annual number of products ingested from external sources by the site. {Maps to requirement 5.4 a}
11. **Products Ingested/Yr per FTE.** The annual products ingested count divided by the total staff effort for the Ingest functional area.
12. **Products Ingested/Yr per Ops FTE**. The annual products ingested count divided by the direct operations staff effort for the Ingest functional area.

4.3.1.2 User Input Parameters

1. **Ingest Product Type Name.** The name of product or data type [add to count of Product Types Ingested/Yr].
2. **External Ingest Interfaces.** The number of distinct external interfaces via which data streams or products are ingested each year.
3. **Ingest Source**. The source or provider of the product or data type. [TBD if this is equivalent to External Ingest Interfaces];
4. **Ingest Delivery Means** . The means of delivery from the source to the data service provider (electronic, media) [TBD];
5. **Ingest Delivery Mode.** The mode of delivery: routine ('real-time' or 'near real time' operational), or ad hoc (intermittent, unscheduled) [TBD];
6. **Products of Type Ingested Per Day.** The typical number of instances (individual products of the type) ingested per day [add to Products Ingested/Yr];
7. **Volume of Type Ingested Per Day.** The average data volume ingested per day for this data or product type [add to Ingest Volume];
8. **Ingest Product Type Format.** Incoming format for product type [add to count of Product Ingest Formats/Yr.];

9. **Conversion Format for Product Type.** The format into which instances of the product type are converted to on ingest, if applicable [for now, add to count of Product Ingest Formats/Yr.];
10. **Ingest Product Type Requirement**. The data service provider's need for this data or product type [Background ??];
11. **Ingest Product Type Retention Period.** The data service provider's planned retention for this data or product type, can be N years after receipt, or indefinite. [use in computing Archive Volume and Archive Products].

4.3.2 Processing

These parameters describe or relate to the generation of products by the data service provider.

4.3.2.1 Internal Computed Parameters

1. **Total Processing FTE.** The total estimated annual FTE effort for the Processing functional area, including any effort in addition to actual operational effort.
2. **Processing Management FTE.** Includes direct management associated with the Processing functional area. Computed from technical and operations staffing.
3. **Processing Technical FTE.** Includes technical staff exclusive of direct operations staff.
4. **Processing Ops FTE**. The estimated annual FTE effort for direct operational activity (e.g. computer operators, production monitors).
5. **Volume of New Products Generated.** The annual volume of new products generated per year by the site. {Maps to requirements 5.5 a, 5.5 b}
6. **Volume of Reprocessed Products Generated.** The annual volume of reprocessed products generated per year by the site. {Maps to requirement 5.5 c}
7. **Processing Volume/Yr.** The annual total volume of new and reprocessed data and/or products that are generated by the site. {Maps to requirements 5.5 a, 5.5 b, 5.5 c}
8. **Processing Volume/Yr per FTE.** The annual processing volume divided by the total staff effort for the Processing functional area.
9. **Processing Volume/Yr per Ops FTE.** The annual processing volume divided by the direct operations staff effort for the Processing functional area.
10. **New Products Generated/Yr.** The annual number of new products generated per year by the site.
11. **Reprocessed Products Gene rated/Yr.** The annual number of reprocessed products generated per year by the site.
12. **Product Types Generated/Yr.** The annual number of different product types generated by the site. {Maps to requirements 5.5 a, 5.5 b, 5.5 c}
13. **Product Generation Formats/Yr.** The number of distinct different product or data formats handled by the Processing functional area.
14. **Products Generated/Yr.** The annual total number of new and reprocessed products generated by the site. {Maps to requirements 5.5 a, 5.5 b. 5.5 c}
15. **Products Generated/Yr per FTE.** The annual products generated count divided by the total staff effort for the Processing functional area.
16. **Products Generated/Yr per Ops FTE**. The annual products generated count divided by the direct operations staff effort for the Processing functional area.

4.3.2.2 User Input Parameters

1. **Product Type Name.** The name of product type [add to count of Product Types Generated/Yr];
2. **Software Source.** A flag that indicates whether the algorithm software produced in-house or received from another activity? [TBD].
3. **QA Function.** A flag that indicates whether the quality assurance is an in-house function or whether another activity involved? [Background];
4. **Production Mode**. Is the generation of this product type performed on demand, or on a routine, scheduled, operational basis? [Background/TBD];
5. **Products of Type Generated per Day.** The typical number of instances (individual products of the type) generated per day [add to Products Generated/Yr];
6. **Volume of Type Generated per Day.** The average data volume generated per day for this product type [add to Volume, Processing];
7. **Product Type Format.** The format in which the new products are produced [add to count of Product Generation Formats/Yr.];
8. **Product Type Requirement.** The data service provider's purpose in generating this product type [Background];
9. **Generated Product Type Retention Period.** The data service provider's planned period of retention for this product type, can be N years after production, or indefinite, or by a rule (e.g. delete if reprocessed) [use in computing Archive Volume and Archive Products].
10. **Reprocessing Capacity for Type.** The data service provider's required reprocessing capacity for this product, as a multiple of the original processing rate [in order to add reprocessing requirement, details TBD]
11. **Reprocessing Plan for Type.** The nominal interval in years at which the data service provider would reprocess the instances of the product type.

4.3.3 Documentation

These parameters describe or relate to the generation, or bringing up to standard, by the data service provider of documentation of data and products, where 'documentation' includes all descriptive information such as catalog metadata as well as user guides, format descriptions, etc.

4.3.3.1 Internal Computed Parameters

1. **Total Documentation FTE.** The total estimated annual FTE effort for the functional area, including any effort in addition to actual operational effort.
2. **Documentation Management FTE.** Includes direct management associated with each functional area. Computed from technical staffing.
3. **Technical FTE.** Includes technical staff working on documentation (including metadata) review, creation, and update.

4.3.3.2 User Input Parameters

None.

4.3.4 Archive

These parameters describe or relate to the archiving of data and products by the data service provider.

4.3.4.1 Internal Computed Parameters

1. **Total Archive FTE.** The total estimated annual FTE effort for the Archive functional area, including any effort in addition to actual operational effort.
2. **Archive Management FTE.** Includes direct management associated with each functional area. Computed from technical and operations staffing.
3. **Archive Technical FTE.** Includes technical staff exclusive of direct operations staff.
4. **Archive Ops FTE**. The estimated annual FTE effort for direct operational activity (e.g. computer operators).
5. **Archive Insert Volume/Yr.** The annual volume of data and/or products that are inserted into the site's archive. {Maps to requirements 5.4 a, 5.5 a, 5.5 b, 5.5 c, 5.7 a}
6. **Archive Insert Volume/Yr per FTE.** The annual Archive Insert Volume divided by the total staff effort for the Archive functional area.
7. **Archive Insert Volume/Yr per Ops FTE.** The annual Archive Insert Volume divided by the direct operations staff effort for the Archive functional area.
8. **Product Types Archived/Yr.** The annual number of different product types added to the site's archive. {Maps to requirement 5.7 a}
9. **Product Archive Formats/Yr.** The number of distinct different product or data formats handled by the Archive functional area.
10. **Products Archived/Yr.** The annual number of products added to the site's archive. {Maps to requirements 5.4 a, 5.5 a, 5.5 b. 5.5 c, 5.7 a}
11. **Products Archived/Yr per FTE.** The annual products archived count divided by the total staff effort for the Archive functional area.
12. **Products Archived/Yr per Ops FTE**. The annual products archived count divided by the direct operations staff effort for the Archive functional area.
13. **Primary Archive Volume.** The year by year cumulative total volume of data contained in the site's primary archive. {maps to requirement 5.7 d}
14. **Backup Fraction.** The fraction of the Primary Archive Volume that is to be backed up. {maps to requirement 5.7 d}
15. **Backup Archive Volume** . The year by year cumulative volume of data contained in the site's backup archive. {maps to requirement 5.7 g}
16. **Archive Volume.** The year by year total cumulative volume of data contained in the site's primary and backup archives. The sum of Primary Archive Volume and Backup Archive Volume.
17. **Archive Volume per FTE.** The archive volume divided by the total effort for the archive functional area.
18. **Archive Volume per Ops FTE.** The archive volume divided by the direct operations staff effort for the Archive functional area.
19. **Archive Media Units**. The number of media units (e.g. tapes) required to hold the data contained in the site's archive.

4.3.4.2 User Input Parameters

1. **Archive Media Type** . The archive media type(s) used by the data service provider. [Background];
2. **Media Standard.** The standard that this media type is in compliance with, or none [Background];

3. **Archive Media Unit Capacity.** The volume of data that can be written to a single unit of the archive media type. [used to compute Archive Media Units];
4. **Archive Media Fill Rate.** The average or typical fraction of a single archive media unit that is filled with archived data or products. [used to compute Archive Media Units];

Note: Have to allow for multiple archive media types. Items 1, 3 and 4 above are used in conjunction with Archive Volume to project Archive Media Units.

5. **Archive Backup Plan**. The data service provider's plan for backing up its archive, including the fraction of the primary archive that is backed up - copied to storage media. [used TBD to contribute to Archive Volume, Archive Media Units];
6. **Archive Migration Plan.** The plan that the data service provider has to migrate its archive to a new media and/or archive system, including the period in years between migrations and the migration rate. [used TBD];
7. **Archive Monitoring.** Archive quality monitoring to support preservation, including the fraction of the archive that is scanned for media integrity per year. [used TBD].

4.3.5 Distribution

These parameters describe or relate to distribution of products to users, either on an operational basis or in response to user requests (a.k.a. 'ad hoc').

4.3.5.1 Internal Computed Parameters

1. **Total Distribution FTE.** The total estimated annual FTE effort for the Distribution functional area, including any effort in addition to actual operational effort.
2. **Distribution Management FTE.** Includes direct management associated with the Distribution functional area. Computed from technical and operations staffing.
3. **Distribution Technical FTE.** Includes technical staff exclusive of direct operations staff.
4. **Distribution Ops FTE**. The estimated annual FTE effort for direct operational activity (e.g. computer operators, distribution technicians).
5. **Distribution Volume/Yr.** The annual volume of data and/or products that are distributed by the site.
6. **Distribution Volume/Yr per FTE.** The annual distribution volume divided by the total staff effort for the Distribution functional area.
7. **Distribution Volume/Yr per Ops FTE.** The annual distribution volume divided by the direct operations staff effort for the Distribution functional area.
8. **Product Types Distributed/Yr.** The annual number of different product types distributed by the site.
9. **Product Distribution Formats/Yr.** The annual number of distinct different product or data formats handled by the Distribution functional area.
10. **Product Types/Yr Distributed Operationally.** The annual number of product types distributed on an operational basis - on a schedule or by rule to specified users.
11. **Product Formats/Yr Operational.** The annual number of different product formats used for products distributed operationally.
12. **Network Products/Yr Operational.** The annual number of network products distributed operationally.
13. **Network Volume/Yr Operational.** The annual volume of data/products distributed by network.

14. **Product Formats/Yr By Request.** The annual number of different product formats distributed by in response to user request.
15. **Products Distributed/Yr.** The annual number of products distributed by the site.
16. **Products Distributed/Yr per FTE.** The annual products distributed count divided by the total staff effort for the Distribution functional area.
17. **Products Distributed/Yr per Ops FTE**. The annual products distributed count divided by the direct operations staff effort for the Distribution functional area.
18. **Internal Catalog Size.** Internal catalog search and order function size - number of product instances included in the catalog.
19. **Network Distribution Volume.** The annual volume of data distributed by the site by network, usually by FTP. {maps to requirement 5.8 f}
20. **Network Distribution Products/Yr.** The annual number of products distributed by the site by network.
21. **Distribution Media Units/Yr**. The annual number of media units (i.e. the sum of the number of tapes of various sorts, CD-ROMs, DVDs, etc., used for distribution by the site). {maps to requirement 5.8 h}
22. **Distribution Media Types/Yr.** The types of distribution media used by the site (CD-ROM, DVD, 8mm tape, etc.).

4.3.5.2 User Input Parameters

1. **Internal Catalog Search Complexity.** The complexity of the search capability offered to the user: 1 - search for instances of single product type by time and space; 2 - search for instances of multiple product types by time and space; 3 - search for instances of product types by geophysical parameter, time, and space across multiple product types.
2. **External Catalog Search and Order.** The type of interface, if any the data service provider provides to an external search and order capability: 1 - none, 2 - external user interface client accesses local catalog information, provides user requests to data service provider, 3 - local catalog information provided to external catalog system which provides user requests to data service provider.
3. **Distribution External Interfaces.** The number of distinct external interfaces, applies to Ingest and Distribution functional areas.

Routine, scheduled, or operational delivery/distribution of products, assumed to be via network, the data service provider provides, including for each product type delivered:

4. **Product Type Name.** Name of product type [add to count of Product Types Distributed/Yr];
5. **Distribution Destination**. Distinct destinations of operational distribution. [add to External Interfaces Count];
6. **Timeliness.** Timeliness requirement, if any [use TBD - germane to level of service];
7. **Delivery Means.** Means of delivery (electronic or media) [use to sort other items to network or media parameters];
8. **Delivery Format.** Delivery format, if converted from local production or archive format [add to count of Product Distribution Formats/Yr];
9. **Products Delivered/Day.** The delivered product count of this type per day, if different from production or ingest information given above. [Add to Distribution Products/Yr, and as applicable, Network Distribution Products/Yr, Media Distribution Products/Yr];

10. **Volume Delivered/Day.** The delivered volume of this type per day if different from production or ingest information given above. [Add to Distribution Volume, and as applicable, Network Distribution Volume, Media Distribution Volume];

Ad hoc, on request delivery or distribution (by network and media) of products the data service provider provides, including:

11. **Users Requesting Products/Yr.** The number of distinct users requesting products per year [add to Users];
12. **User Product Requests/Yr.** The number of product requests received per year.
13. **By Request Products/Yr., Media.** The number of products provided per year, on media in response to user requests [Add to Products Distributed/Yr, and Media Distribution Products/Yr];
14. **By Request Products/Yr., Network**. The number of products provided per year, electronically by network [Add to Distribution Products/Yr and Network Distribution Products/Yr];
15. **By Request Volume/Yr, Media.** The volume of products provided per year in response to user requests on media [Add to Distribution Volume, and Media Distribution Volume];
16. **By Request Volume/Yr, Network.** The volume of products provided per year in response to user requests electronically by network. [Add to Distribution Volume, and Network Distribution Volume];
17. **Distribution Format.** Alternative distribution formats offered by a data service provider, where a conversion is done prior to delivery from the locally generated or stored format [add to Product Formats Handled, Distribution].
18. **Distribution Media Type.** List of types of distribution media used by the data service provider.
19. **Distribution Media Units/Yr by Type.** The number of units per year of each type of distribution media provided by the data service provider.

4.3.6 User Support

These parameters describe or relate to user support provided by the data service provider.

4.3.6.1 Internal Computed Parameters

1. **Total User Support FTE.** The total estimated annual FTE effort for the User Support functional area, including any effort in addition to the direct user support effort.
2. **User Support Management FTE.** Includes direct management associated with the User Support functional area. Computed from technical and operations staffing.
3. **User Support Technical FTE.** Includes technical staff exclusive of direct user support staff.
4. **User Support Ops FTE**. The estimated annual FTE effort for direct user support representatives.
5. **User Contacts/Yr per FTE.** The annual number of user contacts divided by the total effort for user support. Applies to User Support functional area.
6. **User Contacts/Yr per Ops FTE.** The annual number of user contacts divided by the FTE effort for direct user support. Applies to User Support functional area.

4.3.6.2 User Input Parameters

1. **User Level of Support.** A general, TBD as yet, measure or index of the level of user support provided by the site. [TBD].
2. **Users.** The number of distinct users to who contacted user support staff in the course of a year. {maps to requirement 5.9 a}

3. **User Contacts/Yr.** A count of all user contacts - emails, phone calls, etc., handled by the site's user support staff.

4.3.7 Instrument / Mission Operations

These parameters describe or relate to instrument and, if applicable, mission operations functions performed by the data service provider. Instrument monitoring, command generation, event scheduling, etc., is assumed to be a 24x7 activity.

4.3.7.1 Internal Computed Parameters

1. **Total Instrument FTE.** The total estimated annual FTE effort for the Instrument / Mission Operations functional area, including any effort in addition to actual operational effort.
2. **Instrument Management FTE.** Includes direct management associated with the Instrument / Mission Operations functional area. Computed from technical and operations staffing.
3. **Instrument Technical FTE.** Includes technical staff exclusive of direct operations staff.
4. **Instrument Ops FTE**. The estimated total annual FTE effort for platform and instrument operations.
5. **Platform Operations FTE.** Includes monitoring status and performance of, and generate commands for spacecraft. {maps to qualitatively to requirement 5.3 a}
6. **Instrument Operations FTE.** Includes monitoring status and performance of, and generate commands for, instrument(s). {maps to qualitatively to requirement 5.3 a}

4.3.7.2 User Input Parameters

1. **Platforms Monitored.** The number of platforms whose performance, health and safety, etc., are monitored by the data service provider.
2. **Platform Actions per Yr.** The number of platform commands generated for upload, platform events scheduled, etc., per year.
3. **Platform Flag.** Indicates whether or not the data service provider uses the services of a platform operator's mission operations system (e.g. provides commands to a NASA or other operator facility for validation and uploading).
4. **Instruments Monitored.** The number of instruments the data service provider is responsible for monitoring.
5. **Instrument Actions per Yr.** The number of instrument commands generated for upload, instrument events scheduled, et., per year.

4.3.8 Sustaining Engineering

These parameters describe or relate to sustaining engineering (i.e. software maintenance and enhancement of operational software) performed by the data service provider.

4.3.8.1 Internal Computed Parameters

1. **Total Sustaining Engineering FTE.** The total estimated annual FTE effort for the Sustaining Engineering functional area, including any effort in addition to actual operational effort.
2. **Sustaining Engineering Management FTE.** Includes direct management associated with the Sustaining Engineering functional area. Computed from technical staffing.
3. **Sustaining Engineering Technical FTE.** Includes technical staff engaged in software maintenance.

4.3.8.2 User Input Parameters

1. **SLOC Maintained.** The number of lines of code that are maintained by the site, of custom (site developed rather than COTS) software used to support the functional areas. Includes reused software. Maintenance is assumed to be equivalent to sustaining engineering - enhancement as well as bug fixes. Applies to internal support only. {maps to requirement 5.11 a}

4.3.9 Engineering Support

These parameters describe or relate to engineering support provided by the data service provider.

4.3.9.1 Internal Computed Parameters

1. **Total Engineering Support FTE.** The total estimated annual FTE effort for the Engineering Support functional area, including any effort in addition to actual operational effort.
2. **Engineering Support Management FTE.** Includes direct management associated with the Engineering Support functional area. Computed from technical staffing.
3. **Engineering Support FTE.** Includes engineering and technical effort that is not otherwise called out, e.g. system engineering, network engineering, test engineering, system administration, and database administration. Applies to data service provider. {maps qualitatively to requirement 5.2.3 a, 5.2.3 b}

4.3.9.2 User Input Parameters

None.

4.3.10 Technical Coordination

These parameters describe or relate to technical coordination performed by the data service provider.

4.3.10.1 Internal Computed Parameters

1. **Total Technical Coordination FTE.** The total estimated annual FTE effort for the Technical Coordination functional area.
2. **Technical Coordination Management FTE.** Includes direct management associated with the Technical Coordination functional area. Computed from technical staffing.
3. **Technical Coordination FTE.** Includes technical staff directly engaged in technical coordination.

4.3.10.2 User Input Parameters

None.

4.3.11 Implementation

These parameters describe or relate to system implementation performed by the data service provider.

4.3.11.1 Internal Computed Parameters

1. **Total Implementation FTE.** The total annual estimated FTE for the implementation area.
2. **Implementation Management FTE.** Includes direct management associated with implementation.
3. **Software Development FTE.** The total estimated annual FTE for software development, integration, and test, if this is computed by functional area. This will be projected from the amount of software to be developed and the implementation period.
4. **Implementation Engineering FTE.** The estimated annual effort for engineering support to system development, e.g. system integration and test, configuration management.
5. **Custom Software, SLOC.** The size of the software required, if this is computed by functional area. This will be projected from mission parameters that size the system needed.

4.3.11.2 User Input Parameters

1. **Software Reuse Fraction.** The amount of software that will be reused from previous projects. The precise formulation is TBD; it must allow for rework of reused software, etc.

4.3.11.3 Cost Model Output Parameters

1. **Development Staff, FTE.** The annual FTE for development effort, technical excluding management.
2. **Development Staff Cost.** The cost for development staff, using the development staff labor rate.
3. **Hardware Purchase Cost.** The cost for data system hardware needed by the data service provider. This will be projected from mission parameters that size the system needed.
4. **COTS Software Purchase / License.** The cost for purchase of COTS software package and/or annual license costs.
5. **Facility Preparation Cost.** All costs associated with preparation of the facility to house the data service provider, and lease, utilities, etc., during the implementation period.
6. **Total Implementation Period FTE.** The annual sum of all implementation period FTE components.
7. **Total Implementation Period Cost.** The annual sum of all implementation period cost elements.

4.3.12 Management

These parameters describe or relate to management, administrative, and related functions performed by the data service provider.

4.3.12.1 Internal Computed Parameters

1. **Total Management FTE.** The total estimated annual FTE effort for the Management functional area.
2. **Center-Level Management FTE.** Includes center level 'front office' management and administration. Computed from overall functional area staffing.
3. **Functional Area Management FTE.** Includes the sum of the direct management FTE associated with the other functional areas. Computed from functional area management FTE parameters.
4. **Internal Support FTE.** Includes property management, logistics, consumables procurement, facility support, etc., within the data service provider. {maps qualitatively to requirement 5.2.2 d}

4.3.12.2 User Input Parameters

None.

4.3.13 Facility / Infrastructure

These parameters describe or relate to facility support and infrastructure maintenance performed by the data service provider.

4.3.13.1 Internal Computed Parameters

None.

4.3.13.2 User Input Parameters

1. **External Net Connection.** A list of external network connections that the data service provider supports.
2. **Source / Service.** The vendor that is the source of the network connection or provider of the network service.
3. **Bandwidth.** The nominal bandwidth or class of service or capacity of the network connection.

4. **Recurring COTS Software License Cost.** Cost of annual renewal / update of COTS licenses. Placeholder for now!
5. **Facility Area**. The area in square feet required to house the data service provider.
6. **Data System Area**. The area within the facility required to house the data service provider's data system(s).

4.3.13.3 Cost Model Output Parameters

1. **Recurring Network / Communications Cost.** The cost associated with network connectivity required by the data service provider.
2. **Recurring COTS Software Cost.** The cost for COTS upgrades or licenses during the operating period.
3. **Hardware Maintenance Cost.** The annual cost of maintaining the system hardware, assumed to be TBD a fraction of the hardware purchase cost.
4. **Supplies Cost.** The annual cost of supplies, including storage and distribution media.
5. **Recurring Facility Cost.** The total annual facility cost, including lease, utilities, etc., during the operating period.

4.3.14 Site Level Parameters

These parameters describe or relate to the data service provider site as a whole. In some cases they are roll-ups of (selected) functional area parameters listed above.

4.3.14.1 Internal Computed Parameters

None

4.3.14.2 User Input Parameters

None

4.3.14.3 Cost Model Output Parameters

1. **Management Staff, FTE.** The annual FTE associated with management and administration, including financial administration, supervision, and other administrative functions. Includes overall data service provider management as well as management associated with individual functional areas.
2. **Management Staff Cost.** The cost for management staff, above, using the management staff labor rate.
3. **Technical Coordination Staff FTE.** The annual FTE associated with supporting SEEDS technical coordination processes, including developing and maintaining common standards and interfaces.
4. **Technical Coordination Staff Cost.** The cost for technical coordination staff, above, using the technical coordination staff labor rate.
5. **Sustaining Engineering FTE.** The annual FTE associated with sustaining engineering, which includes bug fixes and enhancements to custom software.
6. **Sustaining Engineering Staff Cost.** The cost for sustaining engineering staff, using the sustaining engineering staff labor rate.
7. **Engineering Support FTE.** The annual FTE associated with system engineering, system administration, database administration and other general technical support.
8. **Engineering Support Staff Cost.** The cost for engineering support staff, using the engineering support labor rate.

9. **Operations Staff FTE.** The annual FTE for all aspects of data service provider operations, including system operations, user support, etc.
10. **Operations Staff Cost.** The cost for operations staff, using the operations staff labor rate.
11. **Total Operating FTE.** The annual sum of the operating FTE components.
12. **Total Operating Cost.** The annual sum of all operating cost elements.

4.3.15 Control Parameters

These parameters provide control information for execution of the cost estimation model. Some apply across data service providers, rather than to a particular data service provider.

4.3.15.1 Internal Computed Parameters

None

4.3.15.2 User Input Parameters

1. **Annual Inflation Rate.** The annual rate of inflation to be applied to all recurring staff costs, lease costs, or license costs.
2. **Hardware Discount Rate.** The annual rate at which the cost of hardware of constant capacity is projected to decline.
3. **COTS Software Discount Rate.** The annual rate at which the cost of COTS software of constant capability is projected to decline. User Input.
4. **Implementation Period.** The number of mission years over which development costs are spread - implementation is assumed to start with mission year 1.
5. **Management Staff Labor Rate**. The fully loaded labor rate for management and administration.
6. **Technical Coordination Staff Labor Rate.** The fully loaded labor rate for management and administration.
7. **Development Staff Labor Rate.** The fully loaded labor rate for development staff.
8. **Operations Period.** The number of mission years over which operations costs are spread.
9. **Operations Start.** The mission year during which operations are assumed to start.
10. **Operations Staff Labor Rate.** The fully loaded labor rate for operations staff.
11. **Sustaining Engineering Staff Labor Rate.** The fully loaded labor rate for sustaining engineering.
12. **Engineering Support Labor Rate.** The fully loaded labor rate for engineering support.

4.4 Cost Estimation Model Output Parameters

These parameters, defined in section 4.3 above, are the output that will be produced by the cost estimation model; i.e. they comprise the initial draft of the content of the cost estimate. They correspond to the cost estimate output tables shown in Section 2.2 above. They are grouped into costs (and support information) for the initial implementation period, followed by costs (and support information) for the operations period.

4.4.1 Initial Implementation Period

1. Management Staff, FTE.
2. Management Staff Cost.
3. Technical Coordination Staff FTE.

4. Technical Coordination Staff Cost.
5. Development Staff, FTE.
6. Development Staff Cost.
7. Hardware Purchase Cost.
8. COTS Software Purchase / License.
9. Facility Preparation Cost.
10. Total Implementation FTE.
11. Total Implementation Cost.

4.4.2 Operations Period

1. Management Staff FTE.
2. Management Staff Cost.
3. Technical Coordination Staff FTE.
4. Technical Coordination Staff Cost.
5. Sustaining Engineering FTE.
6. Sustaining Engineering Cost.
7. Engineering Support FTE.
8. Engineering Support Cost.
9. Operations Staff FTE.
10. Operations Staff Cost.
11. Development Staff FTE.
12. Development Staff Cost.
13. Recurring Network / Communications Cost.
14. Recurring COTS Software Cost.
15. Hardware Purchase Cost.
16. Hardware Maintenance Cost.
17. Supplies Cost.
18. Recurring facility Cost.
19. Total Operating FTE.
20. Total Operating Cost.

4.5 Cost Estimation Model User Input Parameters

These parameters, defined in section 4.3 above, must be provided by the user when executing the cost estimation model.

These parameters apply to both implementation and operations. They include control parameters that apply to the data service provider, such as labor rates and planned implementation and operation periods, and parameters that describe the mission workload planned for the data service provider. These mission parameters drive the sizing of the data service provider, and the sizing drives the estimated costs.

4.5.1 Control Parameters

These are overall control parameters that are required for any data service provider whose costs are to be estimated.

1. Annual Inflation Rate.
2. Hardware Discount Rate.
3. COTS Software Discount Rate.
4. Implementation Period.
5. Management Staff Labor Rate.
6. Technical Coordination Staff Labor Rate.
7. Development Staff Labor Rate.
8. Operations Period.
9. Operations Start.
10. Operations Staff Labor Rate.
11. Sustaining Engineering Staff Labor Rate.
12. Engineering Support Labor Rate.

Others TBD.

4.5.2 Mission Parameters

This set of parameters constitutes a complete description of the mission requirements the data service provider must meet, and thus constitutes the sizing information for the data service provider. These parameters are derived from mission descriptions for data service providers. Mission descriptions from actual data service providers will be used to build the comparables database, and mission descriptions for future data service providers will be a source for cost estimation input parameters.

Mission parameters will be listed by functional area in the sections that follow below. Each section will contain a list of the information that will be collected from data service providers for that area. Some of the information is needed for a background understanding of how the data service provider functions and is more directly related to the requirements and levels of service discussed in Section 5.

4.5.2.1 Ingest

Mission parameters for the ingest function are drawn from a description of the data or product streams the data service provider ingests. The description includes the information listed below for each data or product type.

1. Product Type Name.
2. Ingest External Interface
3. Source.
4. Ingest Delivery Means.
5. Ingest Delivery Mode.
6. Products of Type Ingested Per Day.
7. Volume of Type Ingested Per Day.
8. Ingest Product Type Format.
9. Conversion Format for Product Type.

10. Product Type Requirement.
11. Product Type Retention Period.

4.5.2.2 Processing

Mission parameters for the processing function are drawn from a description of the product streams the data service provider generates. The description includes the information listed below for each data or product type.

1. Product Type Name.
2. Software Source.
3. QA Function.
4. Production Mode.
5. Products of Type Generated per Day.
6. Volume of Type Generated per Day.
7. Product Type Format.
8. Product Type Requirement.
9. Product Type Retention Period.
10. Reprocessing Capacity for Type.
11. Reprocessing Plan for Type.

4.5.2.3 Documentation

TBD. Mission parameters for the documentation function are drawn from a description of the product streams the data service provider ingests or generates and adds to its archive. The scope of the documentation can be indicated by a) a count of the product types the data service provider handles, since there can be extensive documentation of each product type, and b) a count of the number of product instances the data service provider handles, since there will be documentation associated with each product instance, if only to identify its unique spatial and temporal coverage. Another dimension is the complexity of the documentation, which may be driven by documentation standards that the data service provider uses on its own accord or is required to use.

4.5.2.4 Archive

Mission parameters for the processing function are drawn from a description of the product streams the data service provider ingests and generates. Details concerning the retention on the archive of data and products ingested by the data service provider from external sources or generated locally by the data service provider are included in the ingest and processing information described above. The required archive capacity can be projected from that information.

1. Archive Media Type.
2. Media Standard.
3. Archive Media Unit Capacity.
4. Archive Media Fill Rate.
5. Archive Backup Plan.
6. Archive Migration Plan.
7. Archive Monitoring.

4.5.2.5 Distribution

Mission parameters for the distribution function are drawn from a description of the catalog search and order and operational and ad hoc distribution services the data service provider provides.

1. Internal Catalog Search Capabilities.
2. External Catalog Search and Order.
3. Distribution External Interfaces.

Routine, scheduled, or operational delivery/distribution of products the data service provider provides, including for each type delivered:

4. Product Type Name.
5. Distribution Destination.
6. Timeliness.
7. Delivery Means.
8. Delivery Format.
9. Products Delivered/Day.
10. Volume Delivered/Day.

Ad hoc, on request delivery or distribution of products the data service provider provides, including:

11. Users Requesting Products/Yr.
12. Product Requests Received/Yr.
13. By Request Products/Yr, Media.
14. By Request Products/Yr, Network.
15. By Request Volume/Yr, Media.
16. By Request Volume/Yr, Network.
17. Distribution Format.
18. Distribution Media Type.
19. Distribution Media Units/Yr by Type.

4.5.2.6 User Support

User support services provided by the data service provider, including:

1. User Contacts Per Year.
2. User Level of Support.

4.5.2.7 Instrument / Mission Operations

Instrument monitoring, command generation, event scheduling, etc., is assumed to be a 24x7 activity.

1. Platforms Monitored.
2. Platform Actions per Yr.
3. Platform Flag.
4. Instruments Monitored.
5. Instrument Actions per Yr.

4.5.2.8 Facility / Infrastructure

These are non-staff items required to support data service provider operations.

1. External Net Connection.
2. Source / Service.
3. Bandwidth.
4. Recurring COTS Software License Cost.
5. Facility Area.
6. Data System Area.

5 Requirements / Levels of Service Analysis

This section presents the first step in the results of the requirements / levels of service analysis, which is development of a requirements / levels of service template for a generic ESE data service provider that is consistent with and linked to the general data service provider reference model described in Section 4. The linkage between the requirements template and the reference model will be the reference model parameters that must be specified to accomplish cost estimation.

The term 'template' is used for two reasons. The first is that all of the requirements / levels of service will not apply to all types of ESE data service provider. The second reason is that the requirements contain placeholders for specifics that must be filled in (i.e. choices between alternatives shown, or between possible levels of service, or replacement of placeholders with lists or numerical values) to generate from the template a set of requirements / levels of service that apply to a specific ESE data service provider, and that allow a cost estimate for it to be produced.

Not all requirements have levels of service associated with them; by their nature, some requirements are either met or not met without any shades of gray. For example, item 5.7 b bebw states: "The data service provider shall provide for secure, permanent storage of data at the "raw" sensor level (NASA Level 0 plus appended calibration and geolocation information)." There is no fuzz on this requirement, it will apply in full force, or not apply at all, depending upon whether or not the data service provider is responsible for that specific form of data.

The second step in the requirements analysis will be the mapping of requirements / levels of service to the ESE data service provider types, presented in Section 6.

Section 5.1 contains a draft set of high level or programmatic requirements referred to as "NewDISS Level 0 Requirements" produced by the SEEDS Formulation Team in September, 2001. These provide an umbrella for the more specific requirements that follow, and downward traceability will be indicated, and, for now, the term 'NewDISS' is retained in this section as this is a reproduction of the original document.

Sections 5.2 through 5.9 present the general requirements / levels of service template that corresponds to the general data service provider reference model. As such it does not imply or embody any architecture, i.e. any allocation of requirements to various particular components. The requirements are organized into categories as site-wide (those that apply to a general ESE data service provider as a whole), and by functional area. Where appropriate, individual requirements are associated with levels of service. Placeholders for items to be specified when the template is to be used to generate requirements for a specific data service provider are enclosed in brackets […]. Upward traceability back to the draft program level requirements will be indicated.

Guidance for the initial set of requirements and levels of service was drawn from the ESDIS Project Level 2 Requirements for EOSDIS Version 0, updated March 2000, which addressed requirements and levels of service, the report "Global Change Science Requirements for Long-Term Archiving", NOAA-NASA and USGCRP Program Office, March 1999, and the report "Ensuring the Climate Record from the NPP and NPOESS Meteorological Satellites", NRC Committee on Earth Studies, September 2000.

5.1 Program Level Requirements

This section contains the set of program level requirements drafted by the SEEDS Formulation Team in September, 2001, as "NewDISS Level 0 Requirements". The cost model requirements template that follows fits within the general framework of the program level requirements in this section. Traceability to the template will be indicated by a reference in parentheses. When the requirement applies generally the traceability will be simply 'general'. For now, where applicability is uncertain, an '?' will be shown.

5.1.1 General Requirements

d. Data service providers will fully participate (TBD) in NewDISS community-based management processes including standards and interface determination, reuse/architecture refinement, metrics collection, and Enterprise peer review. (see 5.2.1 b)

e. All data service providers will comply with NewDISS Level of Service requirements for core functions and data products (TBD) and will adhere to NewDISS required core interfaces and standards (TBD). Deviation from core standards must be requested and approved via the NewDISS waiver process (TBD). (general)

f. Data service providers will provide metrics (TBD) on data production and utilization to the NEWDISS Office on a routine (TBD) basis. (see 5.5 d, 5.7 g)

g. Data service providers and projects will participate in an annual (TBD) broad-based peer review of ESE data management activities. (see 5.2.1 b)

h. ESE Mission Projects will produce a Life Cycle Data Management (LCDM) Plan. Changes to the LCDM plan will be approved by the NEWDISS Office (TBR). (see 5.2.1 b)

i. To the extent possible and where cost effective, data service providers will reuse software and system components developed by previously NASA funded activities. Projects will enable possible reuse of their software available by following the system design guidelines provide by the NEWDISS reference architecture (TBD). (general)

5.1.2 General Science Requirements

a. Data service providers will provide support to and receive technical direction from an appropriate NASA ESE science parameter team.

b. Principal Investigators will propose a suite of standard science products subject to peer review approval of an Algorithm Theoretical Basis Document.

c. Each data service provider will have a Science Advisory Group that will review progress and plans on a routine basis. (see 5.2.1 b)

5.1.3 Production, Archive, and Distribution Requirements

a. All raw data will be acquired will be calibrated and geolocated to a reference sphere. Calibrated and georeferenced data will made available to all users. (see 5.5 a - referred to as Level 1B).

b. Data at the “raw” sensor level (NASA Level 0 plus appended calibration and geolocation information) must be archived permanently. (see 5.6 b)

c. All standard science data (Level 1b, Level 2, and Level 3) produced will be made available to any user who requests it without discrimination. (see 5.7 c)

d. All standard data products available to a science team member will be made available to general science users. (covered by c above?)

e. All standard data produced will be archived until the end of the science mission or until transfer to an approved permanent archive. (see 5.6 c)

f. Data service providers will receive orders for data products from the general public and will fulfill those orders with an average delivery time (elapsed time between when the order was completed and product was shipped) of less than five working days. (see 5.7 d, 5.7 f, and 5.7 g; unclear what time interval referred to is)

5.1.4 Standards and Interface Compliance

a. Metadata for all standard products will be produced in accordance with the NEWDISS core metadata standard (TBD). (see 5.7 a)

b. Metadata for all archived standard data products must be searchable by spatial and temporal extent, and must be locatable by the general user via the world wide web. (see 5.7 d)

c. Standard data products made available to the LTA, to another NEWDISS data service provider and to users will be available in one of the NEWDISS core formats (TBD). (see 5.7 i)

d. All standard data products will be cataloged in the Global Change Master Directory (GCMD). Data service providers will provide Directory Interchange Format (DIF) documents on all standard data products to the GCMD prior to release of the data products. (see 5.7 b)

5.2 Site-Wide Requirements / Levels of Service

These requirements apply to the reference data service provider as a whole.

5.2.1 Management

a. The data service provider shall provide management and administrative staff to perform supervisory, financial administration, and other administrative functions.

b. The data service provider shall provide staff required for participation in SEEDS management processes, architecture refinement, metrics collection, annual Enterprise peer review, development and maintenance of a life cycle data management plan, and support for its Science Advisory Group. (from 5.1.1 a, 5.1.1 e, 5.1.2 c)

c. The data service provider shall provide staff required for participation in SEEDS processes for developing and maintaining common standards and interface definitions. (from 5.1.1 a)

5.2.2 Internal Support

a. The data service provider shall maintain system security and data integrity while providing easy access to its data and information services for its user community.

b. The data service provider shall provide and maintain a fully furnished and equipped, environmentally controlled, physically secure facility to house its staff, systems, and data and information holdings.

c. The data service provider shall provide a backup facility for its data and information holdings.
   Levels of Service:

   1) an environmentally controlled and physically secure off-site backup archive facility;

   2) an on-site but separate environmentally controlled and physically secure off-site backup facility;

   3) a backup capability within the data service provider's primary data system(s).

d. The data service provider shall perform resource planning, logistics, supplies inventory and acquisition, and facility management.

5.2.3 Engineering Support

a. The data service provider shall perform system administration, network administration, database administration, coordination of hardware maintenance by vendors, and other technical functions as required for performance of its mission.
   Levels of Service:

   1) no or very infrequent interruptions of data service provider operations;

   2) occasional interruptions in data service provider operations;

3) as needed, with interruptions in data service provider operations a secondary concern.

b. The data service provider shall perform systems engineering, test engineering, configuration management, COTS procurement, installation of COTS upgrades, network/communications engineering and other engineering functions as required for performance of its mission.

Levels of Service:

1) no or very infrequent interruptions of data service provider operations;

2) occasional interruptions in data service provider operations;

3) as needed, with interruptions in data service provider operations a secondary concern.

5.3 Instrument and Mission Monitoring and Command Requirements

a. The data service provider shall monitor the status and performance of [name] instruments and in some cases also [name] spacecraft for which it is responsible, generating instrument commands and in some cases spacecraft commands as needed.

b. The data service provider shall obtain the services of a NASA (or other spacecraft operator as appropriate) mission operations facility to provide instrument and spacecraft data and to receive, validate, and transmit instrument and/or spacecraft commands to the spacecraft.

5.4 Ingest Requirements / Levels of Service

a. The data service provider shall ingest the following data [ingest data stream table, listing for each data stream: name, source, product types ingested, products ingested per day of each type, volume ingested per day]. The input data streams should cover all data to be received by the center, e.g. satellite data streams, ancillary data products, processed products generated by other data service providers, etc., based on its ESE mission.

Levels of Service:

1) operational (time-critical) ingest with immediate verification of data integrity and quality;

2) routine ingest and verification of data quality and integrity without tight time constraints;

3) ad hoc or intermittent ingest on a non-operational basis with verification of data quality and integrity;

4) ad hoc or intermittent ingest on a non-operational basis.

Levels of service can be mixed within a data service provider; i.e. different levels may be appropriate for different data streams.

5.5 Processing Requirements / Levels of Service

a. The data service provider shall generate the following standard products, included required Level 1B products [standard product table, listing for each product type/series: name, product instances produced per day, volume per day, required input data streams] on a highly reliable, operational basis, either on a routine schedule or on-demand, based on its ESE mission.

Levels of Service:

1) standard products shall be generated within 2 days of ingest/availability of required inputs;

2) standard products shall be generated within 7 days of ingest/availability of required inputs;

3) standard products shall be generated within 30 days of ingest/availability of required inputs.

b. The data service provider shall generate the following products [product table, listing for each product type/series: name, average product instances produced per day, average volume per day, required input data streams] on an ad hoc, non-operational basis.

Levels of Service:

1) specific targets for processing adopted on a case by case basis;

2) general goals for processing;

3) no goals, purely ad hoc processing.

c. The data service provider shall reprocess standard products [standard product table] on an ad hoc basis in response to reprocessing requests.

Levels of Service:

1) the capacity for reprocessing shall be 9 times the original processing rate;

2) the capacity for reprocessing shall be 6 times the original processing rate;

3) the capacity for reprocessing shall be 3 times the original processing rate.

d. The data service provider shall provide standard metrics on production to the SEEDS Office. (from 5.1.1 c)

5.6 Documentation Requirements / Levels of Service

a. The data service provider shall generate and provide standard compliant catalog information (metadata, including browse) and documentation describing all data and information held by the data service provider. (from 5.1.4 a)

Levels of Service:

1) data and product holdings documented to the standard for long term archiving;

2) documentation ensured to be sufficient for current use;

3) documentation only as received from product provider.

b. The data service provider shall update documentation of data and products with user comments.

Levels of Service:

1) data and products routinely updated with user comments;

2) data and products occasionally updated with user comments;

3) data and products rarely updated with user products.

c. The data service provider shall generate and provide DIF (directory interchange format) documents to the Global Change Master Directory on all products available from the data service provider prior to their release for distribution. (from 5.1.4 d)

5.7 Archive Requirements / Levels of Service

a. The data service provider shall add to its archive or working storage the following data and products [archive product table, drawn from ingest data stream table, standard product, and ad hoc product tables and reprocessing volume] and related documentation / metadata.

b. The data service provider shall provide for secure, permanent storage of data at the “raw” sensor level (NASA Level 0 plus appended calibration and geolocation information). (from 5.1.3 b)

c. The data service provider shall provide for secure storage of all standard or other science products it produces until the end of the science mission or until transfer to an approved permanent archive, per its data management plan. (from 5.1.3 e)

d. The data service provider shall provide for an [archive] [working storage] capacity of [number] TB.

Levels of Service:

1) archive capacity is cumulative sum of all data ingested plus all products generated;

2) archive capacity is limited to a specified threshold.

e. The data service provider shall perform quality screening on data entering the archive (e.g. read after write check when data is written to archive media) and exiting the archive (e.g. track read failures and corrected errors or other indication of media degradation on all reads from archive media).

Levels of service:

1) exit and entry screening;

2) entry screening.

f. The data service provider shall take steps to ensure the preservation of data in its archive.

Levels of service:

1) 10% per year random screening;

2) 5% per year random screening;

3) 1% per year random screening.

g. The data service provider shall provide a backup for its [archive] [working storage] (using the backup facility per requirement 5.2.2 c under Internal Support.

Levels of service:

1) full off-site backup, with regular sampling to verify integrity;

2) partial, [Backup Fraction - % of archive backed up], off-site backup, with sampling;

3) partial, [Backup Fraction - % of archive backed up], on-site backup with sampling.

h. The data service provider shall use robust archive media.

Levels of Service:

1) archive media compliant with NARA standards;

2) archive media consistent with commercial practice.

i. The data service provider shall plan and perform periodic migration of archive to new archive media / technology.

Levels of Service:

1) planned migration;

2) no planned migration, but ad hoc migration as need is seen to arise.

(Note - this requirement would not apply to a data service provider with a shorter lifetime than a migration cycle appropriate for its archive media / technology.)

5.8 Distribution Requirements / Levels of Service

a. The data service provider shall provide users with access to all metadata and data and information holdings, including all standard science products (Level 1b, Level 2, and Level 3) produced by the data service provider.

Levels of Service:

1) public access to all users;

2) access to the science community;

3) access to a limited team of scientists.

b. The data service provider shall provide data and products to users in (at a minimum) one of the SEEDS core formats. (from 5.1.4 c)

c. The data service provider shall enhance its distribution capability with supporting services such as subsetting, reformatting, and packaging as needed for performance of its mission.

Levels of Service:

1) supporting services available for most archived data and products;

2) supporting services available for less than half of archived data and products;

3) supporting services available for a few selected data and products only.

d. The data service provider shall provide a world wide web accessible search and order capability to [all users (including the general public) consistent with SEEDS standards and practices; to a limited set of science team members]. (Accessibility consistent with the level of service for requirement 5.8 a above.) (from 5.1.3 f, 5.1.4 b)

Levels of Service:

1) allow search for instances of product types by geophysical parameter, time, and space across multiple product types;

2) allow search for instances of multiple product types by time and space;

3) allow search for instances of single product type by time and space.

e. The data service provider shall provide data to users on an [operational, subscription, and/or in response to request] basis.

f. The data service provider shall perform timely distribution of data and products to users by network, providing an average distribution volume capacity of [number] TB per day. (from 5.1.3 f)

Levels of service:

1) availability of a product for network delivery within ten seconds;

2) availability of a product for network delivery within ten minutes;

3) availability of a product for network delivery within twenty four hours.

g. The data service provider shall perform timely distribution of data and products to users on SEEDS standard media types in response to user requests, providing an average volume capacity of [number] TB per day. (from 5.1.3 f)

Levels of Service:

1) shipping of media product within three days of receipt of request;

2) shipping of media product within one week of receipt of request,

3) shipping of media product within one month of receipt of request.

h. The data service provider shall have the capacity to distribute products on an average of [number] media units per day.

i. The data service provider with final ESE archive responsibility (i.e., a Backbone Data Center unless, for example, a Science Data Service Provider held its products to the time for their transfer to the long term archive) shall transfer its data, products, and documentation (done to the long term archive standard) to the designated long term archive according to its Life Cycle Data Management Plan.

j. The data service provider shall provide SEEDS standard metrics on distribution to the SEEDS Office. (from 5.1.1 c)

5.9 User Support Requirements / Levels of Service

a. The data service provider shall be capable of supporting [number] of distinct, active users per year who request and use data service provider products.

b. The data service provider shall provide a trained user support staff.

Levels of service:

1) one user support staff member per 100 active users;

2) one user support staff member per 500 active users;

3) one user support staff member per 1,000 active users.

(The number of active users is the number of distinct users who request delivery of data and/or information products per year.)

5.10 Development Requirements

a. The data service provider shall design and a data and information system capable of meeting its mission requirements. The design shall address hardware configuration and interfaces and allocation of function to platform. The design shall address software configuration, including COTS, software re-use, and new custom software to be developed, including science software embodying product generation algorithms and/or software facilitating integration of science software provided by outside source(s).

b. The data service provider shall develop a staffing plan that addresses staff required to implement and operate the data service provider over its planned lifetime. The staffing plan shall include a breakdown of positions and skill levels assigned to functions.

c. The data service provider shall develop a facility plan, including planning for space, utilities, furnishings, etc., required to support its staff, data and information system, data storage, etc., and the environmental conditioning to be provided.

d. The data service provider shall accomplish the implementation of its data and information system, including purchase and installation of hardware, purchase or licensing and installation and configuration of COTS software, modification, installation and configuration of re-use software, development of new custom software, and integration of all components into a tested system capable of meeting the data service provider's mission requirements.

e. The data service provider shall provide the staff needed to accomplish all needed in-house development and test activities.

5.11 Sustaining Engineering Requirements

a. The data service provider shall maintain and, as needed, enhance custom software it develops to meet its mission needs, and reused software it customizes and integrates, a total of [number] SLOC.

Levels of Service:

1) no or very infrequent interruptions of data service provider operations;

2) occasional interruptions in data service provider operations;

3) as needed, with interruptions in data service provider operations a secondary concern.

6 ESE Data Service Provider Types

This section describes ESE data service provider types, drawing on the NewDISS concept paper "Draft Version 1.0 - NewDISS: A 6-to-10-year Approach to Data Systems and Services for NASA's Earth Science Enterprise", October 2000 for a starting point. For each data service provider type, this section will present the conceptual description taken from the concept paper and a description of the functions of the data service provider type in terms of the data service provider reference model and its functional areas (which will define the subset of the reference model that applies to the data service provider type). This is followed by tables that map the general requirements / levels of service template presented in Section 5 to the ESE data service provider types. This mapping is the basis for requirements / levels of service templates for each data service provider type.

The NewDISS concept paper introduces its discussion of NewDISS data service provider types: "NASA's ESE has requirements for collection and synthesis of scientific information, for bringing synthesized data products to bear on unanswered scientific questions, and for preserving data and information for future scientific discovery. … NewDISS is therefore seen as consisting of a dynamic network of interconnected components, each responsive to its environment, containing capabilities for change over time through feedback with the science community. These components will be responsible for executing NewDISS data management functions and must allow easy participation by scientists and data and services providers. The components of NewDISS have been conceptualized (October, 2000) as including "Backbone" processing centers, PI-managed Mission Data Centers [here Mission Data Service Providers], Science Data Centers [here Science Data Service Providers], and Multi-Mission Data Centers."

Three additional data service provider types are added:

1. Applications Center, focused on uses and users other than research, given the existence of NASA funded applications activities such as Type III ESIPs and RESACS (see Section 7);

2. Information Center, focused on information describing data and products rather than the data and products themselves, based on discussion at the Formulation Team Retreat, November 7-8, 2001, where 'ECHO' was suggested as a possible future instance, and the GCMD is plainly a currently operational instance.

3. Long Term Archive Center, focused on permanent preservation and archiving of data and products and their documentation and active support to climate research, etc., based on a request from Matt Schwaller, a member of the Formulation Team and leader of the Earth Science Data Life-Cycle study. Long term archiving is strictly speaking not an ESE responsibility, but inclusion of a hypothetical Long Term Archive data service provider type is intended to support planning that NASA is doing with NOAA and USGS, the agencies who have (with NARA) the long term archive responsibility.

A "data service provider" does not necessarily imply a physically distinct institution. An institution such as a NASA center, a university, an organization of another US Government Agency such as USGS or NOAA can host a data service provider or a combination of data service providers. This is equivalent to the existing situation in which the University of Colorado hosts the NSIDC DAAC, or the USGS's EROS Data Center hosts the EDC DAAC.

6.1 Backbone Data Center

This section describes the generic Backbone Data Center.

6.1.1 Backbone Data Center Concept

The following is the concept for Backbone Data Centers, from the NewDISS Concept Paper: "These centers, most likely evolving from some of the current DAAC's, will address NASA's responsibility for preserving and protecting the large volumes of data from the ESE satellite missions. One of the primary roles of the backbone

data centers will be to preserve the basic data. Clearly, NASA can provide a considerable amount of existing infrastructure and technical skill needed to provide satellite mission data downlink and "level 0" or "level 1" data processing. Teaming NASA missions with Backbone Data Centers in the Announcement of Opportunity (AO) process for backup or for generation of basic data products may well be an attractive option for handling some of the core data management requirements of NewDISS. Another role for the Backbone Data Centers will be to acquire products agreed to be scientifically important for preservation and to prepare all these data for long-term archiving. These data centers will need to address network connectivity as part of their on-going activities. Selection of these services will be driven by PI-teaming arrangements, using either NASA-available resources or competitive alternatives. Backbone Data Centers, staffed by professional data managers, provide a core set of historical experience and proven capabilities. As such, they provide a means for risk mitigation against the failure of one or more of the NewDISS components by serving as backup centers for the other parts of the NewDISS. These data centers would most likely be few in number to ensure the cost-effectiveness of the NewDISS."

6.1.2 Backbone Data Center Functions

In general, the Backbone Data Center is expected to provide stable and highly robust services, with a key responsibility for data preservation and documentation, and with a mandate to provide professional data management as a resource for ESE as a whole. Backbone Data Centers are not identified with a particular mission or project but provide data management services in support of multiple missions and the NASA science program in general. Backbone Data Centers have an indefinite lifespan subject to regular review of their performance.

Backbone Data Centers could provide processing functions for NASA missions through teaming arrangements with NASA Principal Investigators, and can serve as a backup to other ESE data service providers.

The paragraphs below will discuss the Backbone Data Center role in each of the general data service provider reference model's functional areas.

Mission and Instrument Command and Control - The Backbone Data Center does not perform this function.

Ingest - The Backbone Data Center performs ingest of a wide variety of data types, ranging from low level data streams to ancillary data to all of the levels of derived products. In some cases the ingest function must be performed on a time critical, operational basis, e.g. for data and supporting information received from operating satellite platforms via NASA or other agency mission operations and communications systems. Quality control on incoming data is critical for lower level (e.g. level 0) data ingested, as the Backbone Data Center must detect bad data and request replacement data from operational sources that may have a limited capability for storing and retransmitting data.

Processing - The Backbone Data Center may perform processing through a teaming arrangement with a flight mission Principal Investigator, which can include large scale (in terms of number of products generated and /or product volume data) operational 'standard product' processing and reprocessing, perhaps with emphasis on Level 1 processing vs higher level derived product processing. Processing by the Backbone Data Center would be highly reliable with tight quality control.

Documentation - The Backbone Data Center ensures that its data and product holdings are documented to the standard for long term archiving, working as necessary with external data sources (e.g. other data service providers) to capture all needed information.

Archive - The Backbone Data Center provides a very robust archive capability, performing insertion of data into archive storage, and preservation of data, metadata, and documentation within the archive. Preservation measures should include quality screening of data entering and exiting the archive, quality screening of archive media, off-site backup with sampling to verify integrity, and accomplishing migrations from one type of media to another.

Distribution - The Backbone Data Center serves a broad user community with a robust search and order and distribution (electronic and media) service, including offering subsetting, reformatting, repackaging in response to user needs. The Backbone Data Center will also transfer data and documentation to designated long term archive centers in accordance with life cycle data management plans.

User Support - The Backbone Data Center provides effective user support for a wide range of users.

Sustaining Engineering - The Backbone Data Center performs sustaining engineering, with no or very infrequent interruption of operational capabilities.

Engineering Support - The Backbone Data Center performs engineering support functions with no or very infrequent interruption of its operations.

Internal Support - Internal support performed by the Backbone Data Center must include supporting an archive facility that is environmentally controlled and physically secure and a separate off-site backup archive.

Management - The Backbone Data Center provides management for its own operation and staff to support its participation in SEEDS system level activities.

6.2 Mission Data Service Provider

This section describes the generic Mission Data Service Provider.

6.2.1 Mission Data Service Provider Concept

The following is the concept for NewDISS Mission Data Service Providers, from the NewDISS Concept Paper: "These data systems are specifically affiliated with instruments or satellite systems. They are either PI-led or facility/project-led. They provide key measurements and standard products from NASA-supported satellite instruments. The key characteristic of the mission data centers [here mission data service providers] is that they will be engineered and implemented as part of an ESE mission proposal. It is anticipated that these Mission Data Centers could leverage the activity at the current ESE data management infrastructure: the ECS flight operations and science data systems and the other hardware and software infrastructure at the DAAC's, the ESIP's, and the SCF's. These data centers will need to address network connectivity as part of their on-going activities. Selection of these services will be driven by PI-teaming arrangements, using either NASA-available resources or competitive alternatives. Mission Data Centers will also need to address satellite/instrument command and control and data downlink. Selection of these services will be driven by PI-teaming arrangements, using either NASA-available resources or competitive alternatives, such as commercially provided or university support services.

Mission Data Service Providers will be responsible for their data management functions during an Earth-observation space flight mission. These data service providers will be funded by the mission selected through the ESE flight programs and will be selected by competitive selection for future ESE missions."

6.2.2 Mission Data Service Provider Functions

In general, the Mission Data Service Provider is an element of a particular ESE mission that exists to provide data management services for the life of that mission. The mission might be might involve an instrument on an independently operated spacecraft (such as SeaWinds on ADEOS) or might include multiple instruments on a dedicated spacecraft (such as Terra or Aqua). The services provided by the Mission Data Service Provider extend from instrument or platform command and control through generation and distribution to mission science team members of science products derived from instrument data for quality assurance, validation, and research. A Mission Data Service Provider would provide instrument data and science products to a Backbone Data Center for distribution to the broad user community and archive after the mission life is completed.

Mission and Instrument Command and Control - The Mission Data Service Provider performs this function for instruments and spacecraft that are part of its mission through NASA or other appropriate operational

mission management services. This includes monitoring instrument and spacecraft performance, generating instrument and (if applicable) spacecraft commands, and event scheduling.

Ingest - The Mission Data Service Provider ingests instrument and spacecraft telemetry and instrument data from NASA or other spacecraft operations and communications systems, and ancillary data needed to support product generation from various sources. Ingest of instrument data and instrument and spacecraft telemetry might be performed on a time critical, operational basis, and the Mission Data Service Provider must detect bad data and request replacement data from operational sources that may have a limited capability for storing and retransmitting data.

Processing - The Mission Data Service Providers will perform small to large scale (in terms of number of products generated and /or product volume data) 'standard product' processing and reprocessing. If the processing is performed to meet the needs of the mission science team only, it can be performed as the team requires. If the processing also must meet the needs of other missions (e.g. as ancillary products), science teams, or other users, it may be performed on an operational basis (especially once processing algorithms become stable). Processing by the Mission Data Service Provider would include tight quality control. The Mission Data Service Provider could team with a Backbone Data Center for processing service, especially if there is a requirement for routine, operational generation of standard products.

Documentation - The Mission Data Service Provider generates complete documentation of its instrument data and all derived products. The Mission Data Service Provider cooperates with a Backbone Data Center that receives its data after completion of its mission to ensure that documentation is brought to long term archiving standards.

Archive - The Mission Data Service Provider would not perform an archive function per se, but would maintain secure working storage of data and products until their transfer to a Backbone Data Center at some time during the mission or after completion of the mission. The Mission Data Service Provider would maintain an off-site back up of all data for which it is responsible, and might use the services of a Backbone Data Center for this purpose.

Distribution - The Mission Data Service Provider provides products to the mission science team for quality assurance, validation, or research, with a search and order capability as needed to meet the needs of the mission science team. The Mission Data Service Provider will also transfer data, products, and documentation to a Backbone Data Center either during its mission as backup or when broader distribution of its data and products is appropriate, or at the conclusion of the mission.

User Support - The Mission Data Service Provider provides close support to member of the mission science team.

Sustaining Engineering - The Mission Data Service Provider performs sustaining engineering, with no or very infrequent interruption of any critical operational capabilities.

Engineering Support - The Mission Data Service Provider performs engineering support functions as needed, but with no or very infrequent interruption of any critical operational capability.

Internal Support - Internal support performed by the Mission Data Service Provider must include supporting a working storage facility that is environmentally controlled and physically secure and a separate off-site backup, for which the Mission Data Service Provider might use the services of a Backbone Data Center.

Management - The Mission Data Service Provider provides management for its own operation and staff to support its participation in SEEDS system level activities.

6.3 Science Data Service Provider

This section describes the generic Science Data Service Provider.

6.3.1 Science Data Service Provider Concept

The following is the concept for NewDISS Science Data Service Provider, from the NewDISS Concept Paper: "These data centers [here data service providers] will collect data from multiple missions for a user community focused on a single research question. There are several examples of these types of Science Data Centers in NASA's Space Science Enterprise. These centers are targeted at specific science questions (perhaps from the NRC Pathways Report) and/or science disciplines, and they directly support research and data analysis for specific research questions. These data centers will address network connectivity as part of their on-going activities. Selection of these services will be driven by PI-teaming arrangements, using either NASA-available resources or competitive alternatives."

6.3.2 Science Data Service Provider Functions

In general, the Science Data Service Provider is a temporary data management capability implemented to support a particular research effort by a limited community of users (which will be called its 'research team'). The Science Data Service Provider operates in a research environment, without the need for robustness and performance as would be the case for an operational environment.

Mission and Instrument Command and Control - None.

Ingest - The Science Data Service Provider obtains data and products required to meet the research objectives of its research team from a variety of sources, including other ESE data service providers, other agency data centers, etc. The ingest would not be performed on a time critical, operational basis.

Processing - The Science Data Service Provider would perform processing, and in some cases reprocessing, of new science products developed by the research team on an ad hoc basis.

Documentation - The Science Data Service Provider generates complete documentation of its science products. The Science Data Service Provider cooperates with a Backbone Data Center that receives its products after completion of its working life (or with the designated long term archive for its products) to ensure that documentation is brought to long term archiving standards.

Archive - The Science Data Service Provider would not perform an archive function per se, but would maintain working storage of products obtained from other sources or science products generated as part of the research effort it supports.

Distribution - The Science Data Service Provider generates complete documentation any new science products developed by the research team that constitute new research quality products to be made available to the general science community (e.g. products cited in publications by members of the research team which should be available other scientists seeking to corroborate or extend the research performed by the team). The Science Data Service Provider will make the products collected to support the research effort readily available to members of the research team, and will perform reformatting, subsetting, or packaging of those products as needed to facilitate their interuse by the research team. The Science Data Service Provider will also transfer new research quality science products and documentation to a Backbone Data Center when broader distribution of those products is appropriate, or at the conclusion of the research effort.

User Support - The Science Data Service Provider provides close support to member of the research team it supports.

Sustaining Engineering - The Science Data Service Provider performs software maintenance as needed.

Engineering Support - The Backbone Data Center performs engineering support functions as needed.

Internal Support - Internal support performed by the Science Data Service Provider includes supporting a working storage facility and a separate off-site backup of any new research quality science products generated by the research effort (e.g. that are cited by research team publications), for which the Science Data Service Provider might use the services of a Backbone Data Center.

Management - The Science Data Service Provider provides management for its own operation and staff to support its participation in SEEDS system level activities.

6.4 Multi-Mission Data Centers

This section describes the generic Multi-Mission Data Center.

6.4.1 Multi-Mission Data Center Concept

The following is the concept for NewDISS Multi-Mission Data Centers, from the NewDISS Concept Paper: "A fourth type of data center [here data service provider] is the Multi-Mission Data Center. An example of the type of data activity to be carried out by such a data center is the generation of consistent time-series geophysical parameters, an activity exemplified by the current National Oceanic and Atmospheric Administration (NOAA)/NASA Pathfinder Datasets program, which is funded by NASA's ESE and carried out by PIs at various institutions. These efforts will take on more importance in the future, since NASA ESE has the requirement for generating time-series of geophysical parameters, while the EOS mission strategy has evolved so that it is now designed to accommodate technological change. Thus, these efforts will include construction of the long-time scale datasets from more than one NASA (or other) mission. These data centers will need to address network connectivity as part of their on-going activities. Selection of these services will be driven by PI-teaming arrangements, using either NASA-available resources or competitive alternatives."

6.4.2 Multi-Mission Data Center Functions

In general, the Multi-Mission Data Center is a temporary but potentially long lived data management capability implemented to support a particular data synthesis effort by a limited community of users (which will be called its 'synthesis team'). An example of a data synthesis effort would be research into how to cross-calibrate and consistently map measurements made by different missions (perhaps overlapping or consecutive) in order to be able to generate a consistent, continuous, long-term, research quality data set spanning multiple instruments/missions, validation of the cross-calibrated data sets, and then the production of the long time series data set. Such a production effort could be quite intensive in order to accomplish in a reasonable time the generation of a long time series data set involve handling many year's worth of a number of good sized data sets. The Multi-Mission Data Center operates in a research environment, without the need for robustness and performance as would be the case for an operational environment.

The distinction drawn between a Science Data Service Provider and a Multi-Mission Data Center is that the former supports a particular research effort, while the latter supports a data synthesis effort that would enable future science efforts using the new, research quality data sets it produces.

Mission and Instrument Command and Control - None.

Ingest - The Multi-Mission Data Center obtains data and products required to meet the research objectives of its synthesis team from a variety of sources, including other ESE data service providers, other agency data centers, etc. The ingest would not be performed on a time critical, operational basis, but could involve large amounts of data if long time series of large data sets are involved.

Processing - The Multi-Mission Data Center would perform processing of new data synthesis products (such as long time series data sets) developed by the synthesis team on an ad hoc basis. This processing could be a major effort, for example if the objective is a long time series product produced from a number of large, multi-year input data sets. The Multi-Mission Data Center could accomplish a large scale processing effort (such as a major effort to generate a long time series data set once the cross-calibration, mapping, etc., involved had been tested and validated) through a partnership with a Backbone Data Center or other processing facility.

Archive - The Multi-Mission Data Center would not perform an archive function per se, but would maintain working storage of data and products obtained from other sources and new data synthesis products generated by the center. This could involve large data volumes, and the working storage would be configured to facilitate the processing effort.

Distribution - The Multi-Mission Data Center generates complete documentation any new data synthesis products developed by the synthesis team that are new research quality products to be made available to the general science community, including full, documentation of the cross-calibration and any other steps taken to build the consistent time series. The Multi-Mission Data Center will make the products collected to support the data synthesis effort readily available to members of the synthesis team. The Multi-Mission Data Center will also transfer new research quality data synthesis products and documentation to a Backbone Data Center when broader distribution of those products is appropriate, or at the conclusion of the data synthesis effort.

User Support - The Multi-Mission Data Center provides close support to member of the synthesis team it supports.

Sustaining Engineering - The Multi-Mission Data Center performs software maintenance as needed.

Engineering Support - The Backbone Data Center performs engineering support functions as needed.

Internal Support - Internal support performed by the Multi-Mission Data Center includes supporting a working storage facility and a separate off-site backup of any new research quality data synthesis products generated by the synthesis effort, for which the Multi-Mission Data Center might use the services of a Backbone Data Center.

Management - The Multi-Mission Data Center provides management for its own operation and staff to support its participation in SEEDS system level activities.

6.5 Applications Center

This section describes the generic Applications Center.

6.5.1 Applications Center Concept

NewDISS Applications Centers will obtain NASA Earth science products and use these, sometimes in conjunction with other Earth science data or any kind of other data to produce special products and/or deliver tailored services to an applications community. These communities could include agriculture, fisheries, urban planning, resource management, many etc., which could derive value from NASA Earth science products if they were suitably formatted or packaged or used in conjunction with other data to produce new products specifically designed to meet the needs of the application community. Examples of existing ESE applications activities include the Socio-Economic Applications Data Center (SEDAC), the Type-III ESIPs, and RESACs, all of which are discussed in Section 7 below.

6.5.2 Applications Center Functions

In general Applications Centers perform the same functions as other data service provider types, the primary distinction being the nature of their user community and therefore their products and services. Thus the Applications Center type actually embraces a wide variety of possible functional models, ranging from large operational activities to small activities providing very focused support to a small user community. Consequently cost modeling of Applications Centers would be as complex as the modeling of all of the other types of ESE data service providers - or, in other words, one could identify a range of sub-types within the spectrum of Applications Centers which would be distinct subsets of a generalized applications center type. This level of complication could be pursued in the future. The best approach to take can be considered once a better understanding of the existing array of applications centers (see Section 7 as it develops) is reached.

Applications Centers can be ESE funded for their lifetime, can be given temporary start-up funding to support their development into self sustaining entities, or can be self funding entities that partner with ESE.

Mission and Instrument Command and Control - None.

Ingest - The Applications Center obtains data and products required as inputs for its applications products from other ESE data service providers, other agency data centers, etc. In some cases the ingest would be

performed on a time critical, operational basis, and in other cases might be on an ad hoc or intermittent basis, and could involve large amounts of data.

Processing - The Applications Center would perform processing of new applications products (such as products for agriculture or fisheries) developed by the Applications Center. This processing could be a major effort if low level data sets of large size are used to generate products on a routine basis.

Documentation - The Applications Center would generate documentation sufficient to support the current use of its products.

Archive - The Applications Center would not be likely to perform an archive function per se, depending perhaps on the commercial value of its products beyond their first use, but would maintain working storage of data and products obtained from other sources and new applications products generated by the center. This could involve large data volumes, and the working storage would be configured to facilitate the processing effort.

Distribution - The Applications Center may distribute its products to either a very limited user community or a very broad user community, operationally or intermittently or on an ad hoc basis depending on its particular mission or business plan.

User Support - The Applications Center provides support to its users according to its particular needs, which will vary considerably from case to case.

Sustaining Engineering - The Applications Center performs software maintenance as needed.

Engineering Support - The Applications Center performs engineering support functions as needed.

Internal Support - The Applications Center performs internal support functions as needed.

Management - The Applications Center provides management for its own operation and staff to support its participation in SEEDS system level activities.

6.6 Information Center

This section describes the generic ESE Information Center.

6.6.1 Information Center Concept

In general the Information Center performs many of the same functions as the Back Bone Data Center, except that the Information Center is concerned with information describing data and products (i.e., one or more types of metadata) rather than the data and products themselves. In general the Information Center will obtain its information from other data service providers, assemble it and make it available to its users, and when its users discover data or products they desire, then help (e.g. by providing links to data service provider websites) those users obtain access to the services of source data service providers.

The addition of this data service provider type was based on discussion at the Formulation Team Retreat, November 7-8, 2001, where 'ECHO' was suggested as a possible future instance. The GCMD is a currently operational instance.

6.6.2 Information Center Functions

The paragraphs below will discuss the Information Center role in each of the general data service provider reference model's functional areas.

Mission and Instrument Command and Control - The Information Center does not perform this function.

Ingest - The Information Center performs ingest of one or more metadata types, ranging from product instance (e.g. granule) level inventory metadata streams to overall product type descriptions or service descriptions. In some cases the ingest function may be performed on a time critical, operational basis, e.g. for inventory metadata received from other data service providers to be posted to the Information Center's inventory on an

operational basis. In other cases, ingest of product type descriptions (etc.) are received on an ad hoc basis and are infrequently updated. Quality control on incoming metadata is critical if the Information Center's database is to be current with consistent and accurate content.

Processing - The Information Center does not perform this function.

Documentation - The Information Center ensures that is own content is consistent and complete but does not generate or maintain any other documentation..

Archive - The Information Center provides only working storage for its database of descriptive information.

Distribution - The Information Center serves a broad user community with a robust metadata search capability. While its 'distribution' is of its own metadata, the Information Center facilitates access to the data and products its metadata describes. This might be in the form of links to source data service provider websites, or the ability to accept a user request for relay to a source data service provider.

User Support - The Information Center provides effective user support for a wide range of users who access its metadata holdings and to the source data service providers who provide the metadata.

Sustaining Engineering - The Information Center performs sustaining engineering, with no or very infrequent interruption of operational capabilities.

Engineering Support - The Information Center performs engineering support functions with no or very infrequent interruption of its operations.

Internal Support - The Information Center performs internal support functions with no or very infrequent interruption of its operations.

Management - The Information Center provides management for its own operation and staff to support its participation in SEEDS system level activities.

6.7 Long Term Archive Center

This section describes the generic Long Term Archive Center.

6.7.1 Long Term Archive Center Concept

The report, "Global Change Science Requirements for Long-Term Archiving" (USGCRP, March 1999), of the results of the science panel that met in a workshop held at NCAR in October, 1998, discussed the essential functions and characteristics of a long term archiving program.

In general the Long Term Archive Center performs most if not all of the same functions as the Backbone Data Center, with the additional focus on permanent preservation and archiving of data and products and their documentation, and active support to climate research, etc., that requires reprocessing of and/or access to long time series of data and products. The Long Term Archive Center participates with ESE data service providers in life cycle data management planning and in a process for obtaining science guidance and priorities for long term archiving.

Long term archiving is strictly speaking not an ESE responsibility, but inclusion of a hypothetical Long Term Archive Center type is intended to support planning that NASA is doing with NOAA and USGS, the agencies who have (with NARA) the long term archive responsibility.

6.7.2 Long Term Archive Center Functions

The paragraphs below will discuss the Long Term Archive Center role in each of the general data service provider reference model's functional areas, drawing on the USGCRP report cited in Section 6.7.1 above. Items in the functional discussion below that are explicitly derived from that report are indicated by an appended '(USGCRP)'.

Mission and Instrument Command and Control - The Long Term Archive Center does not perform this function.

Ingest - The Long Term Archive Center performs ingest of a wide variety of data and product types, ranging from low level data streams to ancillary data to all of the levels of derived products, and their documentation. These products may be new to the center or may be replacements of earlier versions of products already archived by the center.

It is essential that the Long Term Archive Center verify the integrity and quality of data and derived product and associated documentation as it is ingested into the archive (USGCRP).

The ingest would be a transfer from another data service provider, e.g. a Backbone Data Center, according to scenario to be documented in life cycle data management plans. If the transfer is from a research environment (e.g. a Science Data Service Provider) that Long Term Archive Center should proactively reach out to the research source and develop the needed agreements and procedure, assist in planning documentation, etc., (USGCRP). The transfer could be a single bulk delivery, or staged as a series of deliveries over a period of time. The transfer could be by media or network.

Processing - It is essential that the Long Term Archive Center exercise data to produce new products and/or new versions of old products to validate data and product documentation, identify and resolve problems in the data, provide opportunities to scientists within the center to pursue science interests, produce new or updated products that are of value to the science community, provide an opportunity to rethink and reorganize how the data are stored to take into account user access needs as well as accommodate new storage and access technology, and increase data longevity (USGCRP). Typical science processing / reprocessing efforts could include production of long time series of intercalibrated data sets from multiple sources/ sensors to support climate change research.

Processing / reprocessing by the Long Term Archive Center would be on an ad hoc basis, but with tight quality control.

Documentation - It is most essential that the Long Term Archive Center ensure that its data sets and products in the archive are accompanied by complete, comprehensive, and accurate documentation (USGCRP), in accordance with long term archive documentation standard. The center works as necessary with external data sources (e.g. other data service providers) to capture all needed information.

Archive - The Long Term Archive Center provides a very robust archive capability, performing insertion of data into archive storage, and preservation of data, metadata, and documentation within the archive. Preservation and maintenance of data holdings, including ensuring integrity and quality of the data, products, and associated documentation is an essential function of the Long Term Archive Center (USGCRP). Extension of maintenance to include updating of documentation with user comments on the data or product is desirable (USGCRP).

Preservation measures should include quality screening of data entering and exiting the archive, quality screening of archive media, off-site backup with sampling to verify integrity, and accomplishing migrations from one type of media to another. It is essential that the Long Term Archive Center develop and maintain a multi-year data migration plan, and that the center perform integrity checks on archive media between migrations (USGCRP).

Data migrations to new archive technology should be taken as opportunities for processing / reprocessing (USGCRP).

Distribution - The Long Term Archive Center serves a broad user community with a robust search and order and distribution (electronic and media) service, including offering subsetting, reformatting, repackaging in response to user needs. It is essential that the center provide the next and subsequent generation of scientists with appropriate access to, and facilitate their use of, its holdings, where 'access' includes a data set / product search and order function, the ability to deliver data and/or products and supporting information

(documentation) on suitable media or electronically, and choices of format, user options such as subsetting, that facilitate access and use (USGCRP).

User Support - The Long Term Archive Center provides effective user support (a user support staff knowledgeable about the data and products, willing and able to help users identify, obtain, and use the products the need, including making referrals to other sources of data - USGCRP) for a wide range of users.

Sustaining Engineering - The Long Term Archive Center performs sustaining engineering, with no or very infrequent interruption of operational capabilities.

Engineering Support - The Long Term Archive Center performs engineering support functions with no or very infrequent interruption of its operations.

Internal Support - Internal support performed by the Long Term Archive Center must include supporting an archive facility that is environmentally controlled and physically secure and a separate off-site backup archive.

Management - The Long Term Archive Center provides management for its own operation and staff to support its participation in archive related activities. For example, it is essential that the center be actively facilitate the process for deciding which products to include or exclude from, or remove from, the archive (USGCRP). It is essential this process be driven by science priorities and scientific assessments, and that scientists be actively engaged in the process: setting criteria and making decisions (USGCRP). The Long Term Archive Center would participate with the appropriate ESE data service providers in these processes.

6.8 Allocation of Requirements / LOS to Data Service Provider Types

This section presents the mapping of the general template of data service provider requirements and levels of service from Section 5 above to the ESE data service provider types discussed in Sections 6.1 through 6.7 above. This mapping is the basis for separate requirements / levels of service templates for each data service provider type. They in turn become the basis for both building the database of data service provider comparables and for the projection of estimated costs for new ESE data service providers of each type.

Note again that levels of service are associated with specific requirements, and indicate different levels of performance in meeting the requirement, such as meeting a timeliness requirement in 1 day, 10 days, or 20 days. Not every requirement will have levels of service associated with it; as explained in Section 5, by their nature, some requirements are either met or not met without any shades of gray.

The requirements / levels of service templates will vary from data service provider type to data service provider type. The different types of data service provider will not all perform the same functions, and will not all meet the same requirements. Indeed, where different data service provider types do have a requirement in common, different levels of service are often appropriate for different data service provider types. The objective of the mapping is to show which of the general data service provider requirements apply to each data service provider type, and where applicable, to indicate minimum, recommended, and desirable levels of service for each requirement.

The tables in this section are arranged to allow convenient comparison of how the requirements / levels of service apply to the different data service provider types. For convenience, the mapping is broken up into three tables. The first includes the center-wide requirements / levels of service from Section 5.2. The second table includes functional area requirements / levels of service from Sections 5.3 through 5.7, and the third includes the requirements from Sections 5.8 through 5.11.

The first column in each table identifies the requirement group. In the first table, the groups are the subdivisions of Center-Wide requirements from Section 5.2. In the second and third tables, the groups correspond to the functional areas from Sections 5.3 through 5.11.

The second column identifies each requirement by its sub-section number with Section 5.

The third column indicates what levels of service, if any, are defined in the requirement's subsection. For each requirement the column will indicate 'none' if no levels of service were defined, or 1, 2 or 1, 2, 3 or 1, 2, 3, 4 depending on whether one, two, three, or four levels of service were defined.

Each of the fourth through tenth columns contains the requirement by requirement mapping information for one of the ESE data service provider types defined previously in Sections 6.1 through 6.7. For a given data service provider type, the entry in this column represents the predominant weight. For example, a data service provider type may ingest a number of different data streams, and a particular ingest level of service might apply for each one. What is indicated in this table is the ingest level of service that best characterizes the data service provider type, especially for the purpose of cost estimation. A similar example would be the backbone data service provider type, which might perform ad hoc as well as operational processing; in such a case the requirements / levels of service will reflect the operational processing. In any case, when a cost estimate is being made for an actual data service provider, its specific requirements would be used, so that in the previous example its cost estimate would not reflect operational processing if it's mission did not include any.

The following entries are possible for each cell in the data service provider type columns:

Yes - this requirement applies to the data service provider type, no level of service distinction.

N/A - this requirement does not apply to the data service provider type.

M-n - the requirement applies, and the minimum level of service for this requirement for the data service provider type is n. (e.g. M-2 indicates that the minimum level of service is level 2 as given in the requirement description.

R-n - the requirement applies, and the recommended level of service for this requirement for the data service provider type is n.

D-n - the requirement applies, and the desirable level of service for this requirement for the data service provider type is n.

Note that minimum and recommended levels of service may be indicated, or minimum, recommended and desirable levels of service.

Depends - well, this is sort of a TBD for requirements for the Applications Center type, because there can be such a wide variation within this general type.

Table 6-1: Site-Wide Requirements / LOS Mapped to ESE Data Service Provider Types

Center-Wide	Requirement	Levels of Service	Backbone Data Center	Mission Data Service Provider	Science Data Service Provider	Multi-Mission Data Center	Applications Center	Information Center	Long Term Archive Center
Management	5.2.1 a	None	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	5.2.1 b	None	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	5.2.1 c	None	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Internal Support	5.2.2 a	None	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	5.2.2 b	None	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	5.2.2 c	1, 2, 3	M-2, R-1	M-2, R-1	M-3, R-2, D-1	M-3, R-2, D-1	M-3, R-2	M-3, R-2	M-1
	5.2.2 d	None	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Engineering Support	5.2.3 a	1, 2, 3	M-2, R-1	M-2, R-1	M-3, R-2	M-3, R-2	M-3, R-2	M-2, R-1	M-2, R-1
	5.2.3 b	1, 2, 3	M-2, R-1	M-2, R-1	M-3, R-2	M-3, R-2	M-3, R-2	M-2, R-1	M-2, R-1

The next two tables continue the mapping with the functional area requirements / LOS mapped to the ESE data service provider types. There are a few cases where an actual data service provider of given type might not meet a particular requirement contrary to what is indicated in the table. For example, if a Science Data Service Provider provides its data and products to a Backbone Data Center, then the requirement under distribution calling for a data service provider to provide its data, products, and documentation to a Long Term Archive Center (5.8 i) would not apply to that data service provider, and a cost estimate for that Science Data Service Provider would reflect that.

Table 6-2a: Functional Area Requirements / LOS Mapped to ESE Data Service Provider Types

Functional Area	Requirement	Levels of Service	Backbone Data Center	Mission Data Service Provider	Science Data Service Provider	Multi-Mission Data Center	Applications Center	Information Center	Long Term Archive Center
Instrument / Mission	5.3 a	None	N/A	Yes	N/A	N/A	N/A	N/A	N/A
	5.3 b	None	N/A	Yes	N/A	N/A	N/A	N/A	N/A
Ingest	5.4 a	1, 2, 3, 4	M-1	M-1	M-4, R-3	M-4, R-3	Depends	M-2, R-1	M-3, R-2
Processing	5.5 a	1, 2, 3	M-3, R-2, D-1	M-3, R-2, D-1	N/A	N/A	Depends	N/A	N/A
	5.5 b	1, 2, 3	N/A	N/A	M-2, R-1	M-3, R-2	Depends	N/A	M-3, R-2
	5.5 c	1, 2, 3	M-3, R-2, D-1	M-3, R-2, D-1	R-3	N/A	N/A	N/A	M-2, R-1
	5.5 d	None	Yes	Yes	Yes	Yes	N/A	N/A	N/A
Documentation	5.6 a	1, 2, 3	M-1	M-2, R-1	M-2, R-1	M-1	M-3, R-2	N/A	M-1
	5.6 b	1, 2, 3	M-3, R-2, D-1	R-3	R-3	R-3	R-3	N/A	M-2, R-1
	5.6 c	None	Yes	Yes	Yes	Yes	Depends	N/A	Yes
Archive	5.7 a	None	Yes	Yes	Yes	Yes	Depends	N/A	Yes
	5.7 b	None	Yes	Yes	N/A	N/A	N/A	N/A	Yes
	5.7 c	None	Yes	Yes	Yes	Yes	N/A	N/A	N/A
	5.7 d	1, 2	M-1	M-1	M-2, R-1	M-2, R-1	Depends	N/A	M-1
	5.7 e	1, 2	M-2, R-1	N/A	N/A	N/A	N/A	N/A	M-1
	5.7 f	1, 2, 3	M-3, R-2, D-1	N/A	N/A	N/A	N/A	N/A	M-2, R-1
	5.7 g	1, 2, 3	M-1	M-2, R-1	M-3, R-2	M-3, R-2	N/A	M-3	M-1
	5.7 h	1, 2	M-2, R-1	R-2	R-2	R-2	N/A	N/A	M-2, R-1
	5.7 I	1, 2	M-2, R-1	N/A	N/A	N/A	N/A	N/A	R-1

One final table, on the next page, completes the mapping of functional area requirements / levels of service to ESE data service provider types.

Table 6-2b: Functional Area Requirements / LOS Mapped to ESE Data Service Provider Types

Functional Area	Requirement	Levels of Service	Backbone Data Center	Mission Data Service Provider	Science Data Service Provider	Multi-Mission Data Center	Applications Center	Information Center	Long Term Archive Center
Distribution	5.8 a	1, 2, 3	M-1	M-3	M-3, R-2	M-3, R-2	Depends	M-1	M-1
	5.8 b	None	Yes	Yes	Yes	Yes	N/A	N/A	Yes
	5.8 c	1, 2, 3	M-2, R-1	M-3, R-2	M-3, R-2	M-3, R-2	Depends	N/A	M-2, R-1
	5.8 d	1, 2, 3	M-2, R-1	M-3, R-2	M-3, R-2	M-3, R-2	N/A	M-2, R-1	M-2, R-1
	5.8 e	None	Yes	Yes	Yes	Yes	Yes	N/A	Yes
	5.8 f	1, 2, 3	M-2, R-1	M-3, R-2	M-3, R-2	M-3, R-2	Depends	M-1	M-3, R-2
	5.8 g	1, 2, 3	M-2, R-1	M-3, R-2	M-3, R-2	M-3, R-2	Depends	N/A	M-3, R-2
	5.8 h	None	Yes	Yes	Yes	Yes	Yes	N/A	Yes
	5.8 I	None	Yes	Yes	Yes	Yes	N/A	N/A	N/A
	5.8 j	None	Yes	Yes	Yes	Yes	N/A	Yes	N/A
User Support	5.9 a	None	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	5.9 b	1, 2, 3	M-2, R-1	M-1	M-1	M-2, R-1	Depends	M-3	M-3, R-2
Development	5.10 a	None	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	5.10 b	None	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	5.10 c	None	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	5.10 d	None	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	5.10 e	None	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sustaining Engineering	5.11 a	1, 2, 3	M-2, R-1	M-2, R-1	M-3, R-2	M-3, R-2	M-3, R-2	M-2, R-1	M-2, R-1

As indicated above the mapping in these tables would be used to write a set of requirements / levels of service templates, one for each ESE data service provider type. Each template could then be turned into a high level requirements statement for a specific data service provider of its type by filling the items left as placeholders in the template.

7 Existing ESE Data Service Providers and Related Activities

This section describes the currently operating ESE data service providers and some related activities that are relevant to ESE data service provider cost estimation. The emphasis will be on those data service providers and activities that are suitable for use as part of the cost estimation model's database (i.e. the 'comparables' used by the estimation by analogy technique) or as independent test cases for evaluating the performance of the cost estimation model. This section will first describe the existing data service providers and activities grouped by their current classification. The descriptions will focus on the relevance of the data service providers and activities to this study - i.e. their usefulness as "comparables" for the cost estimation model's data base or independent test cases.

The section will conclude with a summary of the mapping of the data service providers and activities to the ESE data service provider types described in Section 6 and indicate which are (candidates in the early going) to be used for the model database or as test cases.

Section 7.1 discusses the EOSDIS Distributed Active Archive Centers (DAACs). Section 7.2 discusses the EOSDIS Science Investigator-led Processing Systems (SIPSs). Section 7.3 discusses the Type 2 Earth Science Information Partners (Type 2 ESIPs). Section 7.4 discusses the Type 3 Earth Science Information Partners (Type 3 ESIPs). Section 7.5 discusses the Regional Earth Science Applications Centers (RESACs). Section 7.6 discusses Pathfinder projects. Section 7.7 discusses NASA non-Earth science (i.e. space science) data service providers or activities relevant to ESE data service provider cost estimation. Section 7.8 discusses non-NASA data service providers that are relevant to ESE cost estimation. Then Section 7.9 summarizes the mapping of the data service providers and activities to the ESE data service provider types described in Section 6 and indicates which ones will be used for the model database or as test cases. Section 7.9 will also consider the possibility that some existing activities may not map to any of the ESE data service provider types, raising the question of whether, assuming that these activities will exist in the future, additional data service provider type(s) may be needed.

As noted, some of the existing data service providers listed in this section will be used to build the cost estimation model's database and to build the independent test data. Data will be collected from these data service providers to be added to the cost estimation model's database or to be used for independent testing of the cost model. The 'raw' data collected from these data service providers will be mapped to the reference model, specifically to the applicable subset of the reference model that applies to the ESE data service provider type that it fits with. This will most likely be documented in an appendix to this report as was done for the Best Practices / Benchmark study, which will include for each center its requirements / levels of service template with the placeholders filled with actual information derived from the input data collected from the data service provider.

7.1 EOSDIS Distributed Active Archive Centers (DAAC's)

The EOSDIS DAACs were established in the early 1990's to provide data management and user services in support of NASA's pre-EOS missions and to prepare for and support NASA's EOS missions, including TRMM, Landsat-7, Terra, and Aqua. The DAACs perform data ingest, processing, archive, and distribution in an operationally robust manner. Processing of EOS mission data is done in combination with the EOS instrument teams who are responsible for the EOS science products and their quality. In some cases the instrument team provides science software to the DAAC for operational production (e.g. MISR aboard Terra). In some cases the processing task is divided between the DAAC and a Science Investigator-led Processing System (SIPS) with the instrument team providing some science software to the DAAC while perform other processing separately itself and providing the generated products to the DAAC (a.k.a. MODIS aboard Terra). In some cases the DAAC and SIPS are a close partnership within a single system framework (e.g. CERES on TRMM and Terra). Finally in some cases all of the processing is done by a separate SIPS and the products

provided to the DAAC (e.g. the TRMM mission except for CERES and LIS). SIPS generally handle distribution of products to the members of the instrument teams. In all cases archive and distribution to the general user community are performed by the DAAC. To date, all of these variations have proved successful.

All eight of the DAACs have developed and are operating their own data systems to perform ingest, processing, archive, and distribution functions, and four of the DAACs are also using the EOSDIS Core System (ECS) to perform those functions for Terra the mission (except for mixed cases at the LaRC and EDC DAACs). The ECS is unique in being a complex system designed to be developed centrally and then distributed to multiple sites to support multiple missions. The SEEDS concept places responsibility for development and implementation of needed data system capabilities with the ESE data service provider as opposed to a central development organization, although ESE data service providers will be free and perhaps encouraged to make use of any existing ECS (and other) data system technology that meets their needs cost effectively. The best parallels to future ESE data service provider implementation and operation will be the independent efforts by the DAACs to implement and operate their own systems, and for that reason only the "non-ECS side" of the four DAACs that have received the ECS will be used in the assembly of the cost estimation model's data base.

All eight DAACs have some form of local search and order capability, and all the holdings of all eight are also accessible through the EOS Data Gateway (EDG), and evolution of the original Version 0 Information Management System, which provides a single user interface to the distributed DAAC catalog information. EDG clients are distributed across the DAACs to provide multiple points of entry for users.

NASA has also designated the DAAC's as near-term Type 1 Earth Science Information Partners (ESIPs), noting that "type-1 ESIPs are responsible for standard data and information products whose production, publishing/distribution, and associated user services require considerable emphasis on reliability and disciplined adherence to schedules" (NewDISS concept document).

7.1.1 Alaska SAR Facility DAAC

The Alaska SAR Facility DAAC (ASF DAAC), hosted by the University of Alaska, Fairbanks, works with SAR (synthetic aperture radar) data collected by international platforms (European, Japanese, and Canadian). The ASF DAAC receives requests for SAR data acquisitions which it forwards to the platform operators. It collects SAR data transmitted from the platforms, catalogs it, and produces some derived products. It maintains an archive of the original data and derived products. The user community of the ASF DAAC is tightly constrained by agreements between NASA and the international spacecraft operators. It includes scientists funded through a NASA A/O for research with SAR data. The ASF data system is not locally developed but a product of the Jet Propulsion Laboratory.

For the purposes of this study, the ASF DAAC is seen as a fair match for any of the defined data service provider types.

7.1.2 EROS Data Center DAAC

The EROS Data Center DAAC, hosted by the USGS's Earth Resources Observation System (EROS) Data Center in Sioux Falls, SD,

7.1.3 Goddard Space Flight Center DAAC

The GSFC DAAC (a.k.a. the Goddard Earth Science DAAC) provides ingest, processing, archive, distribution, and user support services for a wide variety of Earth science data and derived products from NASA missions including UARS, SeaWiFS, Nimbus-7, TOMS, TRMM, and Terra. In most cases the DAAC receives data and science products from instrument team processing facilities, SIPS or SIPS predecessors, including TSDIS (TRMM), the SeaWiFS SIPS, the UARS CDHF, and MODAPS for Terra/MODIS. In the case of MODIS, the DAAC generates a suite Level 1 products and provides these to MODAPS, MODAPS then generates the higher level science products and sends them to the DAAC for archive and distribution. The DAAC developed

a “Version 0” data system for its early work with Nimbus-7, TOMS, UARS and various other data sets, and then added a TRMM Support System (TSS) to archive and distribute data and products collected and produced by the TRMM mission SIPS, TSDIS. The TSS was an evolution from the Version 0 system. The DAAC has produced some data packages containing sets of geophysical data sets mapped to compatible projections to facilitate interuse.

For the purposes of this study, the GSFC DAAC’s Version 0 / TSS side is a good match for a Backbone Data Center type.

7.1.4 Jet Propulsion Laboratory DAAC

The JPL DAAC provides ingest, processing, archive, distribution, and user support services for a variety of oceanographic data and products derived from NASA missions and non-NASA data such as NOAA AVHRR. The NASA missions/instruments supported by the DAAC include SEASAT, NSCAT, QuikScat, SeaWinds and altimetry missions including TOPEX/Poseidon and GEOSAT. The DAAC has worked closely with NSCAT and QuikScat instrument teams, providing processing services and close support to team members. The DAAC has also supported the AVHRR Ocean Pathfinder, generating Pathfinder Sea Surface temperature products from a long term AVHRR data set using algorithms developed by the Pathfinder science team. The DAAC’s data system capability has been locally developed and evolved over the years to meet the needs of the missions supported and the DAAC’s user community.

For the purposes of this study, the JPL DAAC is a good match for a Backbone Center - actually one that also hosts processing functions of a Mission Data Service Provider (e.g. NSCAT, QuikScat, SeaWinds) and Multi-Mission Data Center (e.g. Pathfinder).

7.1.5 Langley Research Center DAAC

The LaRC DAAC (a.k.a. the LaRC Atmospheric Sciences Data Center) provides ingest, processing, archive, distribution, and user support services for a number of NASA missions and instruments, including ERBE, CERES on TRMM and Terra, MISR on Terra, and archive and distribution of products from the ISCCP. There are three seemingly distinct modes in which the DAAC has operated. The DAAC implemented a locally developed “Version 0” system to support its early work with ERBE and ISCCP data and products. Secondly, the DAAC and the CERES instrument team cooperated closely on the local development of LaTIS, the Langley TRMM Information System which supported CERES on the TRMM mission, and as LaTIS was expanded to accommodate CERES on Terra the Atmospheric Data Center was seen as a close partnership of a DAAC and a CERES SIPS, a renaming that did not affect the play in the field. Thirdly, the DAAC received an instance of the ECS, which is being used to support the MISR instrument on Terra. Using the ECS, the DAAC performs operational processing, archive, and distribution of MISR products using science software provided by the MISR instrument team. All three modes of operation have proved successful.

For the purposes of this study, the Version 0 - LaTIS side of the LaRC DAAC is a good match for a Backbone Data Center - actually, in its work with the CERES instrument one that also hosts some functions of a Mission Data Service Provider.

7.1.6 National Snow and Ice Data Center DAAC

The NSIDC DAAC provides ingest, processing, archive, distribution, and user support services for a number of missions and instruments that collect data useful for cryospheric research. This has included SSM/I on a series of DMSP platforms, from which the DAAC has produced long time series of gridded science products. The DAAC archives and distributes Polar Pathfinder data sets which include polar region NOAA AVHRR, NOAA TOVS, NASA SMMR and DMSP SSM/I subsets processed to facilitate interuse. The DAAC is providing archive and distribution for MODIS snow and ice products and will be supporting the forthcoming ICESAT / GLAS mission and two missions that will fly AMSR instruments. The DAAC uses locally developed data system capabilities for all of its work with SSM/I, AVHRR, and related data sets, and uses an instance of the ECS for its work with MODIS snow and ice products.

For the purposes of this study, the 'non-ECS' side of the NSIDC DAAC is a good match for a Backbone Data Center, and one that also hosts the functions of a Multi-Mission Data Center.

7.1.7 Oak Ridge National Laboratory DAAC

The ORNL DAAC provides archive, distribution, and user support for a variety of ground-based and remote-sensed measurements data sets collected by NASA sponsored field campaigns (e.g. BOREAS, FIFE and its follow-on, LBA, OTTER, and SAFARI 2000). The DAAC also holds land validation data (e.g. ACCP, FLUXNET, EOS land validation), and a variety of regional and global data (e.g. climate collections, NPP, vegetation collections, etc.). The DAAC uses locally developed data system capabilities for all of its work, including the Mercury data search system.

For the purposes of this study the ORNL DAAC is a reasonable match for a Multi-Mission Data Center that specializes in field campaign data sets and products and similar climate oriented products that may be characterized as large in number, complex it makeup, and small in individual size. This is a difference from many other data service providers which deal with EOS and other flight mission data that most often include a smaller number of much larger lower level products as well as derived products that are still of good size (although some flight mission data sets are not large, such as SAGE or ACRIM).

7.1.8 Socio-Economic Data and Applications Center

SEDAC's mission is to develop and operate applications that support the integration of socioeconomic and Earth science data and to serve as an "Information Gateway" between the Earth and social sciences. SEDAC collects a variety of socio-economic data and information (e.g. Gridded Population of the World, Global Population Database, Environmental Treaty Texts, World Bank indicators of environmentally sustainable development, World Resource Institute guide to the global environment) and interactive applications (e.g. Model Visualization and Analysis for Integrated Assessment Models of Climate Change). The data and information holdings of the DAAC are a wide variety of individually small items. The DAAC uses locally developed data system capabilities to perform all of its work.

For the purposes of this study, SEDAC is a good functional match for a Applications Center, albeit one that deals with very small data sets, and may not be a good quantitative 'comparable' data point for cost estimation of a Applications Center that works with much larger EOS or other mission data sets.

7.2 EOSDIS Science Investigator-led Processing Systems (SIPSs)

From the NewDISS concept paper, to be replaced or updated: "Under the EOS program, NASA's ESE provided funds for SCF's located at the home institutions of EOS instrument team principal investigators. NASA also provided funds to the Interdisciplinary Science (IDS) Investigators to develop SCF's in their home institutions. In some cases these SCF's are sufficiently robust to allow the EOS principal investigators to generate the higher-level data products for their instruments. These SCF's are designated as Science Investigator-led Processing Systems (SIPS). Examples include MODAPS (MODIS Adaptive Processing System), etc."

7.2.1 TSDIS

The TRMM Science Data and Information System (TSDIS), a science team support facility for the TRMM Project, performs real-time processing and post-processing / reprocessing of TRMM science data. Working with the TRMM principal investigators and science algorithm developers, TSDIS maintains the operational science data processing system and ensures the timely processing of all TRMM science instrument data (from the Precipitation Radar (PR), TRMM Microwave Imager (TMI), and Visible Infrared Scanner (VIRS) instruments). During routine operations, raw instrument data is received in near real-time by TSDIS and then processed by the first tier of science algorithms to produce Level 1 products, calibrated swath-level instrument data. From the Level 1 products, the second tier of algorithms are used to compute Level 2 products, geophysical parameters, such as precipitation rate, also at the swath-level resolution. At the final stage of

TSDIS processing, the third tier algorithms produce Level 3 products, gridded geophysical parameters from the first- and second-tier instrument data.

All TSDIS products are available within 48 hours after input data is received. TSDIS delivers its products to the TRMM instrument team for quality control and monitoring, and handles much of the distribution to TSDIS science community (about 60-80 users). TSDIS delivers its products to the GES DAAC for archive and distribution to the general science community and the public. The GES DAAC also services standing orders to instrument team members.

7.2.2 MODAPS

Under a December, 1999 working agreement with the ESDIS Project, the MODIS Science Investigator-led Processing System (SIPS) was assigned the responsibility for producing Level 2 and Level 3 MODIS products (from Level 1 products produced and provided by the GES (GSFC) DAAC) and deliver MODIS products to the science team for quality analysis and validation, and to the GES, EDC, and NSIDC DAACs for archive and distribution to the general science community and the public.

The MODIS SIPS uses MODAPS (which had its origin as the MODIS Emergency Backup System), to carry out its responsibilities. MODAPS is located in GSFC building 32. MODAPS ingests MODIS Level 1B data, geolocation fields, cloud masks, atmospheric profiles, and ancillary data products from the Goddard DAAC, and generates Level 2 and 3 Land, Ice, Atmospheric and Ocean products and stages them to the ECS Production Data Request (PDR) server at the GES DAAC. The products are then either ingested from the PDR server by the GES DAAC or by the EDC and NSIDC DAACs, according to DAAC product responsibilities. MODAPS also sends products to the MODIS SCFs (Science Computing Facilities) for quality analysis and validation.

MODAPS priorities are set by three key instrument team scientists. MODIS scientists establish production policy guidelines for MODAPS operations.

7.2.3 MOPPITT

7.2.4 SAGE III

7.3 Type 2 Earth Science Information Partners (ESIPs)

From the NewDISS concept paper, to be replaced or updated: “Type 2 ESIPs are responsible for data and information products and services in support of Earth system science (other than those provided by the Type-1 ESIP’s) that are developmental or research in nature, where emphasis on flexibility and creativity is key to meeting the advancing research needs. In addition, NASA’s ESE has awarded technology prototyping funds to the Type 2 ESIP’s with the goal of rapidly developing new methods for exchange of environmental information.”

7.3.1 Distributed Oceanographic Data System

7.3.2 Ocean ESIP

7.3.3 Progressive Mining of Remotely Sensed Data

7.3.4 Global Land Cover Facility

7.3.5 Web-based Terrestrial Environmental Research

7.3.6 GENESIS

7.3.7 Earth System Science Workbench

7.3.8 Seasonal - Interannual ESIP

7.3.9 ESP2Net

7.3.10 SnowSIP

7.3.11 Tropical Rain Forest Information Center

7.3.12 Passive Microwave ESIP

7.4 Type 3 Earth Science Information Partners (ESIPs)

From the NewDISS concept paper, to be replaced or updated: “Type 3 ESIPs are responsible for develop practical applications of earth science data for a broader community. Funding for this group is structured so that there is a partnership between NASA and business or institution that makes up each individual Type 3 ESIP. It is expected that ESIPs of Type 3 will become self-sustaining as a result of the applications developed by their partnership with NASA.”

LIST TBD

7.5 Regional Earth Science Applications Centers (RESACs)

7.5.1 Upper Midwest RESAC

7.5.2 Great Plains RESAC

7.5.3 Southern California Wildfire Hazard Center

7.5.4 Southwest RESAC

7.5.5 Mid-Atlantic RESAC

7.5.6 Nautilus RESAC

7.6 Pathfinder Projects

7.6.1 Ocean Pathfinder

7.6.2 Land Pathfinder

7.6.3 Atmosphere Pathfinder

7.6.4 Polar Pathfinder

The Polar Pathfinder seeks to facilitate the comparison of parameters from different data sets. The SSMR-SSM/I, AVHRR, and TOVS research teams employed a common projection, file naming conventions, and validation conventions to develop consistently processed data sets that are easy for the cryospheric community to combine and contrast.

7.7 NASA Space Science Data service providers and Related Activities

7.7.1 Planetary Data System

7.7.2 Space Telescope Science Institute

Maybe others TBD

7.8 Non-NASA Data service providers

Information on a variety on non-NASA data service providers, both U.S. and international, was collected during the ESDIS sponsored Data Center Best Practices and Benchmark Study. Some of those data service providers are potentially valuable as 'data points' for the SEEDS cost estimation model's data base or as independent test cases.

List TBD, but could include NOAA's Active Archive Center, Eumetsat's MARF/MPEF, CERSAT, BADC, EDC.

7.9 Mapping Existing Data Service Providers to ESE Data Service Provider Types

This section will relate the data service providers and other activities described in the preceding subsections to the ESE data service provider types. This mapping identifies the potential 'comparables' or 'analogs' for the cost estimation by analogy method used by the cost estimation model. Those for which the needed information is available will be identified, and progress towards collection of the information will be reported in updates to this section as the effort proceeds.

8 Cost Estimation Relationships

TBD

The intent for this section is to document the cost estimation relationships used by the cost model. The cost relationships will be derived from the model's data base of information collected from real data service providers, and they will be tested against independent data (i.e. data collected from additional real data service providers that is not included in the model database so that it can be used to test the relationships derived from the model's database). This section, in addition to documenting each cost estimation relationship, would include the results of the independent testing. There will be a cost estimation relationship (or a sequence of them) for each item included in the cost estimation output, as outlined above in Sections 2.2 and 4.3.1 above. This section could take the form of an introduction followed by a series of subsections, one for each cost estimation output parameter, that would define the cost estimation relationship, how it was arrived at, what the error of estimate associated with it is, how it was tested.

October 23:

This is a sketch of a possible approach to an initial, simple version of a cost estimation model that would exercise the cost by comparables approach and perhaps be a jumping off point for the big time cost estimation model to come.

a. Point Cost Estimation Model

The point model is called a 'point' model because it would estimate costs for a data service provider at a single point in time, rather than project a year by year cost profile. The point model would use the existing data, collected previously by the Best Practices / Benchmark study, which is 'point' data - data representative of one 'point' in time, the 'Year 2000 Epoch". This data would be used to construct single point cost estimation relationships. The data service providers for which data is available would be sorted by their functional match to the four defined ESE data service provider types, and a set of cost estimation relationships would be developed for each data service provider type for those elements of cost for which data is available, making some assumptions, to help round out the parameter set.

b. Point and Profile Cost Estimation Model

The goal of the point and profile model would be to take a first, arguably baby, step toward estimating year by year ESE data service provider costs.

The simplifying assumption would be made that each data service provider type's year by year costs would fit a particular rough profile. Then the existing data service provider information could be plotted at one point on the applicable profile according to its state of evolution as of the 'Year 2000 Epoch' for which the data applies. Cost estimates produced by the point estimation model could then be extrapolated backward and forward along the profile (using assumed externalities such as an inflation rate) to yield the year by year cost estimation that is the ultimate goal of this study.

References

1. "NewDISS: A 6-to-10-year Approach to Data Systems and Services for NASA's Earth Science Enterprise - Draft Version 1.0", October 2000.

2. "NewDISS Level 0 Requirements", September 2001, Vanessa Griffin ESDIS/SOO and SEEDS Formulation Team.

3. "ESDIS Project Level 2 Requirements: Volume 5: EOSDIS Version 0, Revision B", March 2000, GSFC.

4. "ESDIS Data Center Best Practices and Benchmark Report", September 2001, SGT Inc.

5. "Ensuring the Climate Record from the NPP and NPOESS Meteorological Satellites", NRC Committee on Earth Studies, September 2000.

6. "Global Change Science Requirements for Long-Term Archiving", NOAA-NASA and USGCRP Program Office, March 1999.

Others TBD.